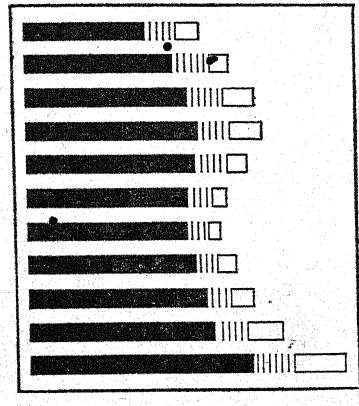


Modern Economics

ARTHUR EDWARD BURNS

GEORGE WASHINGTON UNIVERSITY

ALFRED C. NEAL

DIRECTOR OF RESEARCH
FEDERAL RESERVE BANK, BOSTON

D. S. WATSON

GEORGE WASHINGTON UNIVERSITY

WITH AN INTRODUCTION BY

Howard S. Ellis

UNIVERSITY OF CALIFORNIA

NEW YORK **HARCOURT, BRACE AND COMPANY** 1948

COPYRIGHT, 1948, BY
HARCOURT, BRACE AND COMPANY, INC.

All rights reserved. No part of this book may be reproduced in any form, by mimeograph or any other means, without permission in writing from the publisher.

[c·7·48]

PRINTED IN THE UNITED STATES OF AMERICA

CONTENTS

INTRODUCTION xi

PREFACE xv

PART I THE SCOPE OF ECONOMICS

1 The Bases of Economic Activity 1

2 Change and Growth in the Economy 28

PART II NATIONAL INCOME AND THE EMPLOYMENT OF RESOURCES

3 The Process of Production 57

4 National Income and National Output 84

5 Consumption, Savings, and Investment 106

6 Changing Levels of National Output and National Income 127

7 The Prospects of Economic Progress 151

PART III THE MONETARY PROCESS

8 The Nature and Functions of Money 179

9 The Structure of the American Banking System 203

10 The Functioning of the Banking System 224

11 Monetary Issues and Problems 248

PART IV PRICES AND THE ALLOCATION OF RESOURCES

12 Prices and Price Functions 273

13 Demand 287

14 Supply 313
15 Prices and Markets: I 335
16 Prices and Markets: II 362
17 The Prices of Productive Services 381
18 Wages 405
19 Interest 428
20 Rent 446
21 Profits 459
22 The Price Structure and Price Movements 473
23 Price Control 498

PART V MANAGEMENT AND CONTROL OF ECONOMIC RESOURCES

24 General Aspects of Resource Control 527
25 Business Organization; the Corporation 541
26 Business Finance 559
27 Concentration and Regulation of Industry 585
28 The Supply of Labor and Labor Organizations 605
29 Labor-Management Relations 621
30 Government and the Economic System 647

PART VI GOVERNMENT FINANCE

31 Government Expenditures and Budgets 669
32 The Tax System of the United States 691
33 Tax and Fiscal Policy 709

PART VII INTERNATIONAL ECONOMICS

34 International Trade 735
35 International Finance 753
36 International Economic Policy 785

PART VIII **PERSONAL DISTRIBUTION AND USES OF INCOME**

37 The Personal Distribution of Income 813
38 The Disposition and Use of Income 831
39 The Determinants of Income Distribution 847
40 Governmental Assurance of Income 865

PART IX **TYPES OF ECONOMIC SYSTEMS**

41 Proposals for Reform 893
42 The Economic Systems 916

SELECTED REFERENCES 933

INDEX 937

LIST OF CHARTS

1	Gross National Output, National Income, Income Payments, and Aggregate Demand	101
2	Short-Term Business Expectations	115
3	Long-Term Business Expectations	118
4	Aggregate Demand or Gross National Output by Components	149
5	Personal Income and Its Disposition, Selected Years, 1929-1947	150
6	The Relation Between Utility and Quantity	295
7	The Demand Curve of an Individual	299
8	Changes in Demand	303
9	Elasticity of Demand	306
10	Increase and Decrease in Supply	315
11	Elasticity of Supply	316
12	Price-Cost Relationships of a Firm	322
13	Market Price Under Pure Competition	340
14	Changes in Market Price Under Pure Competition	341
15	Monopoly Price	349
16	Productive Services and National Income	389
17	Diminishing Returns	398
18	Rent on Three Grades of Land	457
19	Consumer Price Index, All Items	481
20	Wholesale Prices, All Commodities	484
21	Wholesale Prices in and After World Wars I and II	514
22	Percentage of World Imports, Exports, and Population, 1938	744
23	Composition of United States Foreign Trade, 1860 and 1937	746

LIST OF TABLES

1	Gainful Workers, 1940	76
2	National Output and National Income, 1929-46	92
3	Sources of National Output, 1941	95
4	National Income by Distributive Shares, 1941	97
5	Disposition of Income	99
6	Percentage Disposition of Income Payments	99
7	Current and Real Output, 1939-42	103
8	Sources of Income Payments	120
9	Composition of Aggregate Demand	130
10	Changes in Income Disposition	133
11	Effects of Increased Investment on Income and Consumer Expenditures	137
12	Volumes of Investment, Given Different Propensities to Consume	141
13	Composition of Aggregate Demand, 1929	145
14	Changes in Aggregate Demand, 1929 and 1933	145
15	Composition of Aggregate Demand, 1937	147
16	Composition of Aggregate Demand, 1945 and 1946	148
17	Currency in Circulation	197
18	Legal Reserves	214
19	Statement of Bank A	225
20	Expansion of Bank Deposits	233
21	Assets and Liabilities of the Federal Reserve Banks	235
22	Effect of Gold Policy on the Money System	258
23	Distribution of Consumer Expenditure, 1935-36	289
24	Demand Schedule of an Individual	298
25	Demand Schedule of a Veteran	300
26	Market Demand Schedule	301
27	Changes in Market Demand	303
28	Demand Schedules of Different Elasticities	304
29	Demand Schedule of Varying Elasticity	308
30	Effect of Increased Income on Purchases	310

31	Demand Schedules of a Commodity in Terms of Prices of a Substitute	311
32	Demand Schedules for a Commodity in Terms of Prices of a Complementary Good	311
33	A Supply Schedule	314
34	Changes in Supply	315
35	Elasticities of Supply	316
36	Cost-Output Data for Firm A	319
37	Costs in Relation to Size of Firm	324
38	Short-Run Supply Schedule	330
39	Demand and Supply: Market Price	339
40	Demand and Supply: Short Run	343
41	Demand for a Product of a Monopolist	347
42	Changes in Weights of Candy Bars	359
43	Returns from Hiring Help at a Parking Lot	397
44	Demand Schedule of a Firm for Labor	411
45	Variations in the Product of Land with Different Combinations of Factors	449
46	Differences in the Yield from Land	451
47	Average Annual Rates of Price Change, 1896-1913	475
48	Relative Price Changes, Four Commodities	478
49	Distribution of Food Expenditures	479
50	Weighted Index Number	480
51	Aggregative Index Number	483
52	Seasonal Movement of Egg Prices	485
53	Average Money and Real Hourly Earnings in U. S. Industries, 1820-1940	488
54	Yield of High Grade Bonds, 1860-1942	490
55	Percentage Decline in Prices and Production in Selected Industries, 1929-33	494
56	Index Numbers of Cost-Price Margins	497
57	Importance of the Corporation	542
58	Securities of Three Corporations	551
59	Assets of Nonfinancial Corporations	553
60	Comparative Consolidated Income Account	573
61	Comparative Consolidated Balance Sheet	575
62	Expenditures of the Federal Government for Selected Years	675
63	Expenditures of State and Local Governments	675

64	Expenditures of State and Local Governments in 1941, by Function	677
65	Tax Burden by Income Groups, 1938-39	693
66	Federal Tax Revenues	695
67	Tax Revenues of State and Local Governments, 1941	702
68	Federal, State, and Local Revenues, for Selected Years, 1920-43	708
69	The Federal Debt, for Selected Years, 1914-47	730
70	Foreign Trade of the United States, for Selected Years, 1800-1940	746
71	United States Exports, by Principal Regions	748
72	United States Exports, Principal Commodities	748
73	United States Imports, by Principal Regions	748
74	United States Imports, Principal Commodities	749
75	United States Balance of Payments in 1946	757
76	Sources of Funds	759
77	Estimated Income Distribution in 1929	823
78	Estimated Income Distribution in 1935-36	825
79	Comparison of Income, Selected Groups	826
80	Proportionate Distribution, All Income Receivers, 1935-36	827
81	Estimated Income Distribution in 1943	828
82	Proportionate Distribution, All Income Receivers, 1943	829
83	Disposition of Family Income, 1929	834
84	Disposition of Family Income, by Income Groups, 1929	834
85	Disposition of Total Income, 1935-36	837
86	Disposition of Income, by Income Groups, 1935-36	837
87	Disposition of Income, by Income Groups, 1935-36 and 1942	839
88	Distribution of Savers and Nonsavers	839
89	Composition of Consumer Expenditure, 1929	842
90	Composition of Consumer Expenditure, by Income Groups, 1929	842
91	Changes in Components of Consumer Expenditure, 1929 and 1935-36	844
92	Estimated Distribution of Wealth, 1927	852
93	Sources of Income, 1930	853
94	Number and Types of Benefits and Average Benefits: Old-Age and Survivors' Insurance, June, 1946	886

INTRODUCTION

Economics is dedicated to the improvement of the material lot of mankind. Every economist worthy of the name has at some decisive and solemn moment sworn a kind of Hippocratic oath to make the world a better place to live in, to break a lance against Want, Disease, Ignorance, and Squalor. In this grand design there is, however, no simple and patent course. The complex forces of the market rarely permit the exact measurements employed by the natural sciences. Much depends upon the acumen, experience, and judgment of the observer; and even with equal endowments, economists sometimes arrive at divergent conclusions. The result is that there exists no officially approved economics, a circumstance for which, in view of its underlying causes, Americans today can be fervently grateful.

Despite its partly "inexact" character, economics has one great and immutable justification: the sincere effort to discover facts and to think about them calmly and objectively is bound in the end to prevail over hearsay, sentimentality, indifference, prejudice, and the mere weight of authority. During its relatively short history of less than two centuries, systematic economics has repeatedly vindicated this faith. Occasionally by brilliant intellectual *tours de force,* but more frequently by the slow processes of time and consensus, satisfactory explanations and illuminating prescriptions of policy have proved their merit in the context of institutions, politics, and technology to which they were applied. Skepticism as to premises, methods, and conclusions meanwhile played as essential a role as positive statements of doctrine.

In this evolutionary process, the writing of systematic expositions of economics for university students has its own special and important functions, despite the curled lip of disdain with which "textbook" is sometimes spoken. First and foremost, the general introductions to economics afford perennial occasions for presenting new developments in the body of thought, pruning out the dead wood, and assimilating the vital elements into a consistent whole. In the contemporary scene, the pressing necessity is a reordering of the whole field of economic thinking to exploit fully the valuable innovations made by Keynes in *The General Theory of Employment, Interest, and Money,* without losing sight of valuable theories antedating this volume.

The authors of the present book belong to the vanguard of writers who are undertaking this function at a level appropriate to a first introduction to economics for undergraduate students. Seizing upon the most significant of Keynes' contributions—the importance of national income in its own right for human welfare and national strength, and as an independent variable of overriding force in most particular markets—they have introduced the student to these concepts at the very outset. They have aspired to orient the multifarious aspects of economic life to this unifying principle throughout.

But they have not permitted this to become a cliché; other valuable innovations both within the Keynesian revolution and aside from it play their part. Indeed, extremists would probably deny that the authors are followers of Keynes in any sense. Were this true, however, it would be difficult to account for new and vital elements in the parts pertaining to monetary and fiscal policy, to interest and wage theory, to inequality of incomes, to the treatment of aggregate employment, and to international trade. Older approaches and techniques have not been lightheartedly discarded: they have been corrected, supplemented, and brought into focus for our modern world by the aid of new ideas.

The new textbook must undertake another role closely allied to the function of systematizing the body of contemporary thought: it must incorporate practical decisions as to which innovations must still be regarded as experimental and hence to be left to the domain of the experts, and which have been sufficiently proved to warrant their transfer to the public. Just as in medical science, there are in economics many "new and unofficial remedies." The writers of the present book have, in my judgment, quite advisedly refrained from displaying a new technique merely for the sake of novelty, if its superiority remains open to question. Thus the "kinked demand curve" may seem a promising device where sellers are few, but it still awaits empirical verification. Similarly in the field of demand theory, the writers have found that they can operate successfully with the familiar Marshallian concepts without recourse to indifference curve technique. Examples could be multiplied but the test is the same: does the accepted technique leave something unexplained, or does the new approach add genuine substance?

Finally, the authors of a general introduction to economics must test all the component parts for their communicability to the ordinary citizen. It is difficult to imagine that any piece of economic analysis can assume much importance in the world of affairs unless it can be couched in language understandable by the layman. The authors of the present book have taken especial pains to write lucidly, directly, and pleasingly, and not to disdain

some repetition for the sake of the beginner. They have resisted the temptation to play over the heads of their audience for the edification of their professional colleagues.

Most laymen, including university students, approach the subject of economics for the first time with lively interest, with numerous ideas of their own as to how the great economic ills of our times can be remedied, and with not a few of their own intense prejudices and misconceptions. All of this can be turned to good account by skillful authors; they will be careful not to smother these sparks under an avalanche of technical terms, geometric and other mathematical presentations, or choke them out by humdrum or obscure writing.

To the American university students into whose hands this book may come, the authors hope to communicate their own belief that economics brings a hopeful message: that the powers of reason and tolerance will prevail over force and ignorance, not only in the affairs of the market place but finally in those higher things which make up life itself.

HOWARD S. ELLIS

Berkeley, California
January 28, 1948

PREFACE

We believe that the beginning study of economics can be effectively guided and integrated by use of the national income approach. That the national income approach is important for economic analysis is widely recognized, yet until now it has not been adapted to a full-length principles book. Its pedagogical merits have been verified for us repeatedly by our classroom experience with the materials of this book.

The student obtains a clear understanding of economic processes when national income analysis is combined with the theory of price and distribution. Because national income analysis easily engages the interest of the student it comes first in order of presentation; throughout the book it is related to the analysis of prices and of income distribution and to money, banking, governmental expenditure, taxation, and international trade. A different sequence from this—starting with the theory of the firm and proceeding to the theory of prices—has some claim to preference, of course. But the significance of prices is not obvious to the beginner, whereas the aggregative analysis is readily comprehensible to him, and is realistic in terms of what he may know of present-day problems.

In still another way we seek to meet the needs of the student. The efforts of many teachers to convey a genuine understanding of the concepts of economics are hampered if the book at hand uses techniques of analysis that are unnecessarily refined. Consequently we eschew technique for technique's sake. The essentials of analysis are necessarily presented here in detail, but though the materials in this book are difficult they are not beyond the capabilities of the college student who gives them careful attention.

Economic problems and policies are treated together with economic principles. Obviously there are many controversial issues in economics, some old and some new. On controversies involving the prospects and means of attaining economic progress we present several main doctrines, and indicate for the student what technical equipment he should apply in his own study of them. This is not a striving for colorless neutrality; honesty compels us to question doctrinaire views, whoever holds them. However, on certain national issues we state our position directly. Our position is

determined by the logic of economic analysis as we conceive it and by our moderately conservative but, we hope, non-doctrinaire political bias. All important decisions of national economic policy are fundamentally political in their motivation as well as in their consequences. Therefore, in our view, economic policy cannot be evaluated only in economic terms; for this nation it must be appraised in the light of the political traditions of a free society.

We wish to acknowledge our indebtedness to Professor Howard S. Ellis of the University of California, who read the book in manuscript and suggested improvements at many places. Professor Yale Brozen of Northwestern University and Dr. Philip D. Bradley of Harvard University also gave us helpful criticism. Yet the approach, the plan of the book, the sequence of treatment, the emphasis and levels and possible errors of analysis, are our sole responsibilities. While the book is a joint product, primary responsibility for the contents of the chapters is as follows: Burns, Chapters 1-7, 24, 34-40, and sections in Chapters 29 and 30; Neal, Chapters 17-23, 25-29; Watson, Chapters 8-16, 30-33, and 41; and Watson and Burns, Chapter 42. Chapters 25 and 26 were drafted by Professor Everett H. Johnson of George Washington University, and Chapters 28 and 29 by Professor Phillip Taft of Brown University.

A selected reading list covering the main subjects of the book follows Chapter 42.

A. E. B.
A. C. N.
D. S. W.

January, 1948

PART I

THE SCOPE OF ECONOMICS

CHAPTER 1

The Bases of Economic Activity

In these times almost everyone regards himself as something of an economist. Indeed, the magnitude and severity of economic problems in recent decades have brought economic issues to the forefront in much popular discussion. The recurrence of prosperity and depression, of general labor shortages in some years, and of heavy unemployment in others, together with the existence of high taxes and of an enormous public debt, has focused wide attention on these problems and on the various proposals for their solution. More immediate, of course, is the fact that no one can avoid having some interest in his own income, and in taxes and the cost of living as they affect him. Similarly no businessman is without opinion on prices, costs, the productivity of labor, and governmental policies as they apply to his operations.

But while the ordinary person is likely to regard the problems of economics as separate issues, as they may affect him from time to time, it is important to recognize that there is an intimate interrelation between the elements making up the study of economics. For example, total employment is affected by—among other things—the volume of government expenditures and the volume of exports; the national debt affects taxes and the operations of banks; prices and wages influence one another, and all these in turn affect the volume of production and national income. The very existence of economic life presupposes the organization and therefore interdependence of many elements. This is not always obvious.

The knowledge of economics likely to be gained by the newspaper reader

does not carry him far, or always in the right direction. Systematic and prolonged study is required to trace the connections and relations which comprise the vastly complicated modern economic systems. Then, too, many economic processes operate unobtrusively. Although, for example, the roles of prices, taxes, savings, and investment in the economic system are of decisive importance, they can be fully demonstrated only by systematic study. Thus while some knowledge of economics is gained inevitably from daily experience it requires expansion and correction, and it must be integrated and systematized by a disciplined study of economics as a science.

As a branch of knowledge economics describes and analyzes the environment in which man works and receives and spends income, as he attempts to secure for himself and his dependents the requisites of material well-being. These activities are of primary interest to the individual: How much can he earn? What are his prospects of greater earnings in the future? Can he save enough for future needs? What occupation or business should he enter? He is concerned with prices and living costs and with taxation—with any policy or circumstance or institution which may affect his material well-being.

The economic system in modern times has become greatly complex. The individual when working, earning, saving, and spending performs some small part of the system's operation. While his productive activity adds something to the flow of commodities and services produced by the system, it may be of no immediate use to him personally; e.g., he may produce a small part that fits into a larger good, or legal advice which he does not need himself, or cotton which is useful only when made into cloth. What he produces is useful only when it is combined and exchanged with other products in the vast system in which he works. Because he *specializes* in producing what he does not want, or in producing more of something than he wants, he must work for money—a wage or salary or other income. With this money he buys the things that thousands of others have helped produce, just as they may buy what he has produced. Specialization of productive activity in the economic system provides no alternative to the making and spending of *money income* as a means of securing the requisites of material well-being. For this reason the welfare of the ordinary person is inevitably dependent on the performance of the economic system as a whole—on the volume and types of goods produced and the level of national income.

WANTS, SCARCITY, AND ECONOMIC ACTIVITY

Whether the economic process is found in the simple form of a primitive community or in the intricate patterns of the modern industrialized nations, it arises from the *scarcities of things in relation to human wants.*[1] Scarcity necessitates the spending of effort and time in the production of commodities and services for the satisfaction of wants. Whether a person produces things directly for his own use, or whether he works to obtain a money income, the basic problem is the same: time and effort must be expended to satisfy wants. Human wants, scarcity, and work are the bases upon which is erected the whole variegated structure of economic activities. The major economic institutions in the economy of the United States—such as the money and banking system, marketing and transportation, the price system, business corporations and labor unions—exist as methods of organizing the production of the things that satisfy human wants.

The definition of economics reflects the fact of scarcity. Economics is usually defined as the study of man's activities in using scarce means (or resources) to satisfy wants. The word "economics" goes back to the early Greek *oikos,* a house, and *nemein,* to manage, meaning "household management." The problem of managing a household (or estate) arises only when its resources are not adequate for all purposes; only then must resources be conserved for what are deemed the most essential purposes. The modern usages are true to the original meaning. It was customary in the eighteenth century to speak of *political economy,* which meant the practice of managing the resources of a people or country. An *economical* person is one who manages or uses his money and other resources with care and frugality, i.e., he *economizes* in order to get the most out of his limited means. The *economy* of a country is a term referring to the organization of the country's resources for the production of commodities and services. Thus, the original term and its modern derivatives all stress the underlying fact of scarcity and the consequent need to conserve and manage resources. Productive effort is organized in such a way as to select from among the innumerable human wants those that can be best satisfied with the limited means at hand. The best satisfaction must be regarded as the *maximum possible satisfaction;* therefore commodities must be produced and in proper proportions which are most preferred by individuals, taking into

[1] Scarcity is a *relation.* Things are scarce in an economic sense when there is not enough of them at hand to satisfy *all* wants for them. For example, the physical existence of untold billions of tons of coal deep in the earth does not mean abundance in an *economic* sense.

account their differences in tastes or preferences and in their means of making wants felt. Knight states that "The commonly recognized function of economic organization is to utilize the limited resources at the command of the social group in bringing about the largest possible satisfaction of the wants of its members." [2] The modern form of economic organization performs this function largely by the payment of money incomes for the use of economic resources (human effort, capital, and natural resources) to produce commodities and services; in turn these incomes are used by those who receive them for the satisfaction of wants.

Human Wants

Wants differ considerably from one time to another and from one social group or individual to another. Some human wants are, of course, elemental and common to all peoples, mainly the wants for food, shelter, and clothing. For a large part of the world's population the effort to keep body and soul together is so difficult that their mental horizons are limited largely to these wants. But even these elemental wants are not simple; people do not want food, shelter, and clothing merely for physical survival. These simple wants exhibit a great variety. In poor and primitive communities, as in those more fortunate, the wants are for certain *kinds* of food, not food in an abstract sense; and the same can be said of other elemental wants of man.

As a social group advances beyond the point where physical survival is the first consideration, their wants exhibit the most varied character. The satisfaction of some wants is a physiological necessity but there are many ways to satisfy them. This diversity is governed by religion, custom, national characteristics as well as by the natural environment.

Convention or custom govern to a large degree individual behavior and social relationships; patterns of wants are imposed on the individual by the social group through the perpetuation of conventions. Consider for example some of the major events in the life of an average American. The birth of a child is attended by certain rituals: the male parent is expected to present cigars to his friends; flowers, gifts, and cards are expected of friends and relatives. Gifts and parties usually follow year after year to commemorate the event. Upon graduation from high school and college the son or daughter is given still other tokens of approval or esteem. The marriage convention prescribes a wedding ring, a trousseau, a reception, a wedding trip, and other ceremonial accompaniments. Conventional wants

[2] F. H. Knight, *The Ethics of Competition*, Harper, 1935, p. 102.

are expressed at the observance of Christmas, New Year, and Easter. Indeed, convention affects in detail the daily wants of the individual. In matters of food, dress, recreation, house furnishings, and even in his relations with family and friends, neighbors, and strangers, the community sets the general pattern of behavior to which each person is expected to conform.

Habit is a second potent influence governing human wants. The ease and convenience of following habit in the selection of cigarettes, brands of food, cosmetics, and the innumerable other items that satisfy recurring and relatively inexpensive wants undoubtedly account for much of the influence of habit in determining wants. If a conscious weighing of choices—if the "real" merits of this brand of cigarette against that—were to replace the role of habit in the satisfaction of daily wants, the cost in time and mental energy to the individual would indeed be enormous. This is not to deny a place to rational choice in the selection of things to satisfy wants, or in the cultivation of wants. Habits are frequently formed on the basis of an original rational calculation of alternative products or methods.

Conventions or customs influence an individual's habitual behavior, yet habit and custom as used here are separable influences on human wants. Convention may impose a pattern of wants on the individual, and subsequently some of these wants become matters of habit. The observance of native customs in a foreign environment is an example. By and large, convention and habit exert a stabilizing influence on wants, for both the group and the individual. They account for a consistency in social behavior; largely because of them there is a marked difference in the wants of various national and cultural groups.

Emulation is a third influence on human wants. The desire to emulate others appears to be common to all persons, but it is particularly important as a determinant of wants in the American society. Lower income groups seek to obtain the same or similar things enjoyed by higher income groups, or those who are considered to be higher in the social scale. The attempt to "follow the leader" in ostentation and display is assiduously promoted by advertising as well as by industry in producing cheap copies of the expensive things bought by the social and financial upper strata. Esteem and envy are large motives in this inequal competition. The desire of those on a low level to emulate the wants of those above them, and the desire of the latter to remain different from those below, inevitably lead to a diversification and expansion of human wants. The desire to emulate others tends to inject instability but also growth into the pattern of wants and into the productive activities designed to satisfy wants.

It is frequently said that convention, habit, and the desire to emulate or

to be different are irrational influences which thwart the maximum satisfaction of wants. For example, custom requires men on special occasions to suffer the discomforts of formal attire, or on warm days the nuisance of a necktie, etc. Habit makes one a slave to certain wants. The desire to emulate leads to an endless and frustrating striving for new things, many of which may be shoddy imitations to be cast away soon for still other things. The desire to be different means only a waste of money, for wants are forcibly and constantly shifted to new objects. But do these influences necessarily lessen want satisfaction?

Obviously, it is often rational for the individual to conform to social conventions, however "irrational" the total effect or origin of this convention may appear to be. And habit has its conveniences. The conscious weighing of choice might cause a person to spend so much time in satisfying certain wants that time for additional leisure or earning power would be lost. The irrationality of emulation is perhaps a moot issue. If a person gets more satisfaction out of attempting to overtake his betters (in want satisfaction) than in not doing so, can this be said to be irrational behavior? The question raises points of morality and social standards which involve more than economic factors. All in all, the charge that convention, habit, emulation, and ostentation are irrational elements in the determination of wants presupposes that the critic has established canons of rationality; but in turn such canons may seem irrational to others.

Wants express the aims, ends, or purposes of individuals. Within the elastic bounds of law and convention the individual is accorded discretion to formulate and pursue his own wants. Some persons want as much leisure as possible, others want an abundance of material goods ranging from the ordinary to the esoteric. While some may seek intellectual attainment, others are content with the ordinary comforts they have habitually enjoyed, there being a virtually infinite variety of personal conceptions of the "good life." These aims or ends, whatever they may be, guide the individual and the group in economic activity—in the earning of money incomes and the use of these incomes.

The Dynamic Character of Wants

Human wants in this society appear to be indefinitely expansible; as goods multiply, so do wants. Modern man does not want merely shelter but a comfortable home and a place for display; not food alone but a varied and well-prepared menu; not raiment to hide his nakedness but a wardrobe suitable for many purposes and occasions and in the latest fashion.

He wants, too, a wide range of amusements, perhaps educational and cultural improvement, and the conveniences and luxuries of modern civilization as they constantly change and multiply. A man may strive to attain a certain standard of want satisfaction (standard of living) for himself and family; but its attainment is often merely a stage in the pursuit of a still higher standard, where wants are more varied and their satisfaction more costly.

The dynamic character of wants is apparent. But this fact does not imply that discontent is rampant. Some people are satisfied with a given range of wants, a given plane of life, and want only to have this week what they had last. Stability of this kind is not, however, characteristic of American society as a whole. Instead, a ceaseless urge to expand wants and to acquire more income to satisfy these wants is a feature of American life. Standards of living deemed adequate in one generation are held inadequate for the following one. In spite of a prodigious increase in material wealth over the past 150 years, more is written on the problem of poverty now than a century or more ago. What is commonly regarded as "material progress" is actually the multiplication of wants faster than the means for their satisfaction.

This ceaseless striving for something more is a deeply ingrained cultural trait. The desire for financial success is virtually unchallenged in this country. Parental duty requires that each generation be admonished to "better itself," i.e., to increase its wants and its capacity (income) to satisfy these wants. Media that mold the public mind—radio, press, movies, magazines, advertisements—build up wants by fostering at once an emulation of others and a desire for distinction and difference. Changing fashions in dress, new models in automobiles, in household fixtures, and in countless other things create dissatisfaction with present possessions and foster a desire for new and different ones.

Perpetual scarcity is the outcome of this proliferation of human wants. Nowhere is nature so lavish with her resources that even the elemental wants are satisfied without effort. The multiplication and diversity of wants beyond this demands still greater effort and ingenuity of man in his attempts to satisfy them.

Wants and Economic Activity

Human wants and productive activity are related in complex ways. Wants guide productive activity; things not wanted will not be produced for long. Aside from the urge for survival, how do human wants arise? Why

are such things as radios, automobiles, electric shavers, and air transport wanted? In an historical sense such wants for specific types of goods are formed after these things make their first appearance; i.e., the *economic process produces both new wants and things to satisfy wants.* Once wants are established, production is directed toward their satisfaction. However, it is generally true that the new want which is created is merely a new variety of some earlier and basic want which is inadequately satisfied once the new good becomes available; e.g., the electric shaver replaces a straight or safety razor.

What leads producers to venture into want-creating activities? Why produce something for which no present want exists? Part of the answer undoubtedly is the hope of profit in the production of something new. However, there is more involved here than this. Productive activity in the modern era is in large measure an end in itself; it is a game or a rivalry. That wants are created and satisfied may be incidental to the basic urge to activity per se, an urge which in an earlier period sent men off to the Crusades, and later to explore new lands, climb mountains, and fly airplanes. When this urge to activity takes the form of producing things that people may learn to want and buy, the producer is doubly paid. He derives satisfaction from the productive activity itself—the building up of a new organization, surmounting difficulties, outwitting his rivals—and he gets a money income to support himself and finance his further activities. Productive activity may thus be a means of self-expression and creative achievement, as well as a means of creating and satisfying consumers' desires. For this reason power and prestige as ends are important in fashioning the character of the productive process and therefore of wants.[3]

Free Goods and Economic Goods

Things that satisfy wants fall into two large categories: (1) free goods and (2) economic goods. Both of these are objects of human desire. They have *utility,* which is their capacity to satisfy wants.

Those useful things that are available freely and abundantly are termed *free goods.* Under most circumstances air is a free good; water from a spring is another. In newly opened territory land may be so abundant that it is a free good. Free goods are not scarce, no want for them goes unsatisfied, and no effort or money is required to obtain them. There is no economic problem connected with these things and they have no place in the

[3] See Knight, *op. cit.*, Chaps. 1 and 2.

consideration of economics (except as they may aid in the production of economic goods).

Economic activity is directed toward the acquisition and use of scarce things, or *economic goods,* comprising both physical commodities and services. Foodstuffs, houses, shoes, radios, books, flashlight bulbs, machine tools, golf clubs, the services of doctors, lawyers, teachers, professional ballplayers—these and countless other things are scarce and therefore known as "economic goods." A thing is scarce when the want for it cannot be fully satisfied; something is sacrificed in order to enjoy its satisfaction. Either the individual's effort is spent directly to acquire it, or money is spent to purchase it.

Some economic goods are not the result of human effort. Land in its natural state, minerals in the earth, a waterfall, and timber in the forests are products of nature which are economic goods when they are too scarce to satisfy all wants for them.

Scarcity, then, is the definitive feature of an economic good; and the manifestation of scarcity is either the expenditure of effort or of money to obtain the good. The term *scarcity,* however, is often used carelessly and wrongly in everyday discussion. For example, during wartime the "scarcity" of butter, bacon, domestic help, shipyard labor, and many other commodities and services was contrasted to prewar "abundance" of these goods. Actually things were merely more scarce during wartime than in the earlier years, for at no time were butter, bacon, domestic help, and other things "free." The same fallacy is committed by those who glibly advance the "economy of abundance" idea. If taken literally the phrase is a contradiction in terms; if abundance prevails no economy is needed. In the same way low prices and ease of purchase are confused with abundance; but the fact that things are produced and sold at a price, even though low, is evidence of scarcity: they are economic, not free, goods.

The scarce *means* of satisfying wants fall into two general categories. First are the economic goods capable of satisfying wants *directly.* These are termed *consumer goods,* commodities such as a loaf of bread or a fur coat, and services rendered directly to consumers, such as those of a night-club entertainer or a physician. Second are those economic goods that satisfy wants *indirectly,* termed *producer goods.* A stamping press in an automobile factory satisfies no consumer wants directly, but by forming fenders and bodies it contributes to the eventual making of finished automobiles which do satisfy consumer wants. The same can be said of a steel mill, an iron-ore mine, a retail delivery truck, and countless other things.

The Limitation of Economic Resources

The commodities and services that satisfy human wants directly are scarce because the economic resources—human effort, capital, and natural resources—that produce them are limited too. Herein lies the central problem of economics: how to use these resources in the production of things for which there are virtually unlimited wants. Wartime poses this problem clearly. At every turn military and civilian wants are confronted by acute shortages—of finished goods and of the labor and other resources to produce them. But war does not raise a new problem in this respect; it merely intensifies an inherent problem.

Economic resources are (1) natural resources, (2) man's own physical efforts and talents, and (3) capital, the man-made aids to production.

The limitation of natural resources is clearly evident. For all practical purposes, land is fixed in amount; mineral resources, too, are fixed in supply and moreover are subject to depletion as they are used; forest resources at best grow but slowly and they also may suffer serious depletion. There have been times when the land and all its wealth seemed boundless—it could be had for the taking. At such times natural resources were abundant and free. This was the condition in the early days of settlement on this continent. But as population grew the era of abundance passed; land and its resources became the property of persons and government, and natural resources soon became scarce in relation to the demand for them.

Man's efforts and talents are of course limited in both quantity and quality. Nowhere are they available for work in the productive process without the promise of remuneration, even if that be no more than enough to sustain life.

Capital—the tools, implements, machines, buildings, installations and other man-made aids to production—is limited because it must be produced. Although more will be said about capital and its various meanings later, a working definition of it is necessary at this point. Capital may be defined as produced goods designed not to satisfy wants directly but to aid in the production of other goods and services; in this sense capital is comprised of producer or instrumental goods. Thus mechanics' tools, farm implements, industrial machinery, a cotton textile mill, a power plant, or raw materials do not of themselves satisfy wants, but enable other things to be produced that can perform this function.

The production of capital involves a *saving*, or setting aside, of part of total productive output; some wants are postponed or sacrificed entirely

so that capital may be produced. This point is of central importance. Since wants are virtually unlimited, man could devote his time and effort exclusively to producing things that satisfy his wants immediately. Instead, part of his time and effort goes into the production of capital, into things that cannot in themselves satisfy wants. In this sense, then, capital involves saving or a deferring of consumption; conversely, saving in this sense comprises that part of productive output not currently consumed. Capital, therefore, is necessarily limited in amount because (1) it must be produced, and (2) it must be saved out of a total production which itself is limited in relation to human wants.

Economics, it has been said, is the study of man's activities in using these limited resources for the production of things to satisfy his many wants. But it is apparent that at any time some labor and capital may stand idle; in times of depression idle workers may number into the millions as factories and mills are shut down. Can it be said that labor is scarce when unemployed men are clamoring at the gates of closed factories? Are not closed factories and idle machines evidence that capital is "abundant"? These depression phenomena in no way affect the validity of the definition. Men and machines do not become idle because there are no wants for the things they can produce. On the contrary, unsatisfied wants increase in such periods because the lack of income prevents many people from buying the things they would otherwise purchase. The idleness does not result because there is more labor and capital than needed, but because the complex mechanism of producing, distributing, and consuming becomes disordered and prevents the full use of these resources.

The Economic Struggle: Economic Costs

Man must struggle to live. The limitation of economic resources leaves no alternative. In its elemental sense man's struggle is to wrest from the natural environment the economic goods that sustain life. The production of man-made economic goods involves a cost. The cost of production may be viewed in two ways: human cost and money expenditure.

Human cost is the time, toil, and trouble required for the production of things; it is measured by the alternative given up in producing one thing rather than something else.[4] Human costs are of two kinds, those connected with *labor* and those with *waiting*. Labor costs involve physical and mental

[4] That is, when time and effort are spent in producing commodity A, all other things that could be produced with this time and effort are given up. The cost of A, then, is measured by the alternative or opportunity forsaken.

fatigue, discomfort, discipline, and other exactions of work. Waiting is a less obvious human cost; it arises from the fact that people generally prefer to have things now rather than later.[5] Wants are usually immediate and insistent, but nevertheless production of the things to satisfy wants requires time; thus there is a frustration of some wants for a time, and this frustration is a cost. A person may prefer to work harder and shorten the period of waiting, or perhaps to wait longer and work less intensely. Both labor and waiting are therefore human costs; and the satisfaction of wants by producing economic goods requires the bearing of these costs.

Money costs of producing things are more tangible and familiar. These include payments for wages, salaries, interest on borrowed money, rent of property, payments for materials, freight, insurance, and other expenses. Some of these money costs are payments to persons who participate directly in producing things; wages, salaries, fees, commissions are of this kind. Profits are payments received by the enterpriser when his business is successful. Rent and interest are payments for the use of assets—land, buildings, and money. When these assets are privately owned, payments for their use become the incomes of their owners. All these costs are *income payments* and must be paid in order to induce people to work and to offer their money and property for the production of economic goods. Stated differently, income payments or money costs obtain the *services* of people and things essential to production.

Costs of production—either human or money—have a common origin in the fact of scarcity. Nature in her niggardliness imposes a scarcity, and man compounds this scarcity by the multiplication of his wants. The economic struggle appears never-ending. In a strict sense "abundance" can never be attained, for it would mean that all wants are satisfied. During the last 150 years each succeeding generation in this country has produced a larger and more varied output than the preceding one. But has scarcity abated? The clamor for higher wages, for more farm income, for higher standards of living by all groups apparently never ceases. A growth in total production of want-satisfying goods merely whets the appetite for more. In man's economic struggle, his adversary—scarcity—never tires.

Scarcity brings with it another struggle. The limited supplies of want-satisfying goods cause a competition among men—among groups and nations—for their shares of the total. Each person's share depends principally upon his money income—his capacity to buy what he needs from the total

[5] The fact of waiting underlies the creation of capital; while capital is being produced, satisfaction of consumer wants is unavoidably deferred. Stated another way, saving involves waiting.

stock of want-satisfying goods. More income for one may mean less for another. Higher wages for labor might reduce the profits of the employer; high profits in one industry might be obtained at the expense of another; the prosperity of one country might rest on its exploitation of colonial areas. It is this aspect of scarcity, the conflict of groups over income, that raises many of the most vexing problems in society.

The Basic Economic Functions or Processes

Because the resources of an economy are limited, economic activities must be organized in some fashion. System, order, or direction are essential to the operation of an economy so that the economic resources can be employed most advantageously and thereby contribute to the satisfaction of as many wants as possible. The organization of the means of want satisfaction is always appropriate to the wants of the people in question. By American standards the economy of a Mexican village would be thought "inefficient," but such a judgment is irrelevant to the point at issue. The village economy is an orderly process by which the inhabitants satisfy wants. Part of the effort or resources of the village is devoted to growing food—many different kinds of food—part to handicraft work, another part to such local government as it deems essential, and so on through the range of economic activities. If more resources were devoted to food production less would be given to satisfying wants for other things, including leisure. The activities of the community become balanced in the course of time in such a fashion as to satisfy tolerably well the more urgent of the wants of the community *as the community itself conceives them.*

All communities and nations are compelled by the scarcity of economic resources to establish systematic patterns of economic behavior in order that no urgent wants go wholly unsatisfied. Of the United States it is sometimes said that the economic system is chaotic and disorganized; taken literally this is nonsense. No established economy is chaotic. The means of organizing its resources may obviously invite improvement, and the emphasis it places on the satisfaction of certain wants against other wants may raise social and ethical questions; but the fact remains that human effort, capital, and natural resources must of necessity be organized, directed, or guided in some way—or no economy exists.

There are, then, certain basic economic functions having a common origin in the scarcity of economic resources. These functions are: (1) *the production of an appropriate combination of goods and services (the output of the economy)*; (2) *the exchange of this output*; (3) *the distribution of*

income among the members of the economy; (4) *the consumption or use of the output.* From what has already been said it may be seen that these functions are unavoidable, whether the economy be a feudal village, a capitalist America, or a socialist Russia. It is the business of economics to investigate these functions in order to determine how an economic system works. In the remainder of this chapter these economic processes will be described briefly in order to summarize the subject matter of this book, and to bring out the cohesiveness of the economic processes.

THE PRODUCTIVE PROCESS: NATIONAL OUTPUT AND NATIONAL INCOME

In the productive process human effort, capital, and natural resources combine to turn out a flow of commodities and services which comprise the *national output.* The size of the national output is one of the principal matters to be examined in this book. The size or, as it is commonly termed, the *level* of national output, and the level of employment of resources in its production are discussed mainly in Chapters 4, 5, and 6. The level of national output and employment is but one aspect of an investigation of the productive process. Fully as important is the study of the *composition* of output; i.e., the amounts and types of things comprising national output and the types of employment required for its production. Chapters 12 through 21 deal with this problem specifically.

National output is defined as the *money value of finished commodities and the money value of services rendered directly to persons.*[6] First, finished commodities are those in their final form for use; consumer goods on retail shelves or in salesrooms, as well as machinery, equipment, and buildings, are representative finished commodities. Their value is the market or sales prices of these commodities. Second, services counted in the measurement and definition of national output are those rendered directly to consumers (or final users), such as the services of physicians, teachers, lawyers who give legal advice to persons (not to business firms), government employees, insurance salesmen, barbers, etc. It should be noted that services are distinguished between those rendered *directly* to persons, and services employed in the production, transportation, and sale of commodities. The latter services of factory employees, transport workers, retail clerks, and all others engaged in producing commodities appear in the prices of the finished products. On the other hand, services rendered di-

[6] National output is defined more precisely in Chap. 4; meanwhile the term is used in the general sense as defined above.

rectly to persons are not a part of commodity production (e.g., medical advice and attention rendered by a physician) and therefore are a distinct component of the national output. These also are measured by the cost or price of the service rendered.

Usually the market values of finished commodities and services rendered directly can be computed so as to obtain annual totals for the national output. Thus the output may be perhaps 150, or 175, or 200 billion dollars in a certain year. Such a total is the end result of the employment of the nation's economic resources during a year; in like manner, changes in the totals from one year to another usually represent changes in the work performed (or employment) by these resources. National output is never large enough to satisfy all wants for the commodities and services it comprises. Even when the efforts and talents of a people and their capital are fully employed the resulting output falls short of satisfying all wants for consumer goods and services and producer goods.

Along with the flow of national output goes a flow of income in the form of wages, salaries, profits, rent, interest, fees, and commissions to those who aid in the production of commodities and services. As noted previously, these are the principal money costs of producing the commodities and services comprising the national output. The income earned in the process of production is termed the *national income.*[7] It too is measured in money, so that in any one year the national income may be perhaps 150, or 175, or 200 billion dollars. The process of production, therefore, provides both the commodities and services for the satisfaction of wants, and the money income which enables people to buy these things.

The connection between income and production should be clearly understood. Income created by the productive process may be viewed as (1) real income and (2) money income. *Real income* comprises the actual commodities and services making up the national output; by this definition real income and the national output are synonymous. *Money income* comprises the wages, salaries, and other payments earned by the tens of millions of income receivers in the economy; it is the principal money cost of producing the real income or national output.

The Level of Economic Activity

In this country, and indeed in all the major communities of the world, there is little dispute over the desirability of maintaining high levels of

[7] National income is defined more precisely in Chap. 4; meanwhile it is used in the general sense defined above.

national output, national income, and employment. The material well-being of a nation is dependent in large measure on the level of its economic activity. For this reason, the causes and conditions governing the level of national output in the immediate future and over the long period of time have long been the subject of economic investigation.

In the United States there are about 60 million persons who make up the labor supply, i.e., persons who normally seek work for a money income. There is, too, a vast amount of productive plant, installations, machines, equipment, and other forms of capital at hand to aid this labor force in the production of commodities and services. Now the employment of both labor and capital is subject to rapid and sometimes drastic changes. At times practically all the labor supply is at work and the factories, mills, and other places of work are actively engaged in production: this may be termed a condition of *relatively full employment.* At other times, many persons are out of work and part of the country's capital facilities lie idle. Such changes are common in this country. Prosperity and high levels of employment are followed by depression and large-scale unemployment, bankruptcies, and severe individual and social losses.

Over a period covering many decades, the economic resources of nations change. This has been so in the past and will continue to characterize the future. The growth in population and labor supply, the accumulation of capital, and the growing efficiency of capital through technological change open up the prospect of long-continued economic progress. But these long-term changes depend, ultimately, on the direction and nature of short-term changes in output, income, and employment.

The size of the national output, national income, and employment depends immediately upon the *decisions of those who employ resources to produce commodities and services.* What prompts these decisions? Under a capitalist system it is chiefly the expectation of producers that they will earn a money income in their productive activities. The production of national output entails a cost; this cost is principally the sum of national income paid and accrued to all who participate in its production. Now it follows from this that the *national income received by the public at large must be returned to producers in the purchase of their output.* Only in this way can producers as a whole recover the cost of producing national output. *Aggregate demand* (i.e., aggregate expenditures) for commodities and services must therefore equal the value or cost of national output, if this output is to be maintained.

Aggregate demand is determined by:

1. The expenditures of consumers—consumer decisions to buy.
2. The expenditure of investors for new capital—investment decisions.
3. The expenditures of government—governmental decisions.

The total income of the tens of millions of income receivers, however, is not spent by them entirely. First, some of it is set aside as savings; another part is taken by government as personal income taxes; both amounts are *not* spent by income receivers on commodities and services for consumption purposes. Consumer expenditures thus fall short of the value or cost of national output. If, on the one hand, savings are borrowed by investors for the purchase of new capital, and on the other hand, if taxes are spent by government for its various public purposes, the whole of the income paid out in the course of producing national output is returned to the productive process. Of these three factors the amount of investment is most crucial, for investment is normally the principal means of putting savings back into productive channels. A fall in investment expenditure usually means a fall in aggregate demand [8] and a consequent decline in national output, income, and employment; a rise in investment usually is accompanied by an increase in output, income, and employment.

Consumer expenditures, too, may initiate changes in total output. Consumers may spend more and save less, or spend less and save more, from one period to another. Government may also change its rate of expenditure. Such decisions effect changes in the size of the national output and national income. This is seen clearly when rapidly expanding governmental expenditures are made for war purposes; or again when, following a war, consumers increase their expenditure by using wartime savings.

The central problems in the analysis of the level of output and income thus hinge on investment and savings. Investment in capital is governed largely by long-term expectations of profit; money is spent on buildings, machines, and equipment in the hope that a profit will be returned over a period of years. These expenditures on capital are financed in effect by drawing on the current savings of the economy. In this way savings are put back into the channels of industry and trade. When the economic resources of a nation are fully employed, national income is high and so are savings.

It is at this point that considerable controversy is found. Why is invest-

[8] Unless a fall in investment is accompanied by an increase in consumer spending through a decrease in savings.

ment periodically inadequate? Will investment in the decades ahead be large enough to absorb the savings that would be set aside under conditions of relatively full employment? What governmental policy is likely to be most conducive to adequate levels of investment? Should savings be reduced gradually (and consumer expenditure increased accordingly) so as to permit high levels of output and income with relatively less investment? Or should savings be unimpaired and investment encouraged so as to achieve the maximum long-term expansion of national output? These and similar questions are greeted with various answers by economists, politicians, labor leaders, and businessmen. The purpose of Chapters 4, 5, 6, and 7 is less to answer these questions than to provide the analytical basis for the student to appraise the problems involved.

THE PROCESS OF EXCHANGE

The second basic function essential to all economic systems is the exchange of the commodities and services produced among the members of the economy. The importance of this exchange function was noted long ago by Aristotle, and of course long before his time exchange or trade had been highly developed. This function is necessitated by *specialization* in production. In the modern period each person and each productive unit—a farm, factory, shop, or office—works on one thing, or a few things, or merely on a part that fits into a larger product. The output of each is destined to be exchanged; modern economic systems largely produce for the market, for sale or exchange at a price.

An automobile plant, for example, specializes in the assembly of cars which are sold to dealers, who in turn specialize in the sale of the product to the public. The assembly plant obtains thousands of pieces and parts from hundreds of other plants specializing in the making of one or a few of these components. In turn, each of these plants obtains materials and parts from many other suppliers. Thus, when the production history of the finished automobile is traced it is found that thousands of different specialized producing units have played a part in its eventual assembly. Each unit is a link in a complex chain of exchanges. The entire productive process is linked together in this way, with the chain of exchanges bringing finished commodities to the final consumer.

In either a highly specialized economy such as the United States or a simple agricultural economy, the need to exchange commodities and services requires a *basis* of exchange and a *means* of exchange. The basis of exchange is *value*, or *price*. The means of exchange is *money*. Two wholly

dissimilar things, such as a bushel of wheat and a work shirt, must be reckoned on a common basis before an exchange can take place. If both are priced at $2 the basis of exchange is established. However, it is unlikely that the owners of wheat and of the work shirt would be able to exchange these products in kind conveniently, even though both were valued equally in the market. In practice each exchanges his product for money, the means or medium of exchange, and with the money he is free to buy the other product or any product available for exchange.

The bulk of the national output is produced for exchange, i.e., for sale in the markets of the country. Its monetary value is its aggregate sale price, and its exchange is effected by the use of money.

The Monetary System

Money is used in virtually every step in the process of producing the national output, which is carried out by innumerable specialized production units. Commodities and services flow from these units to the ultimate buyers of the output. Production requires money payments from producers to those who aid them in production in the form of wages, salaries, rent, interest, and other incomes. These incomes flow back to the productive units when income receivers spend money to buy goods; if part of income is saved it flows back to producers when investment is made. Thus, money is the medium of exchange in the modern economy. Producers exchange money in return for goods sold to them and services rendered to them by employees and others in the productive process. Those who receive money income in turn exchange most of it with producers for the purchase of consumer goods. In the American economy and in all other advanced economic systems, production, income distribution, and consumption are therefore dependent upon exchange and the use of money.

The principal source of money supply is the banking system. Its role as the creator or producer of money is of utmost importance because the basic economic processes rest on the use of money. Too little money or too much money tends to disrupt these processes. It is consequently necessary to understand the methods by which the supply of money is made available—the laws, banking regulations, and business customs and practices, as they operate within the structure of the banking system. Increasingly the conditions governing the supply of money are a concern of government and have become subject to governmental control, in an attempt to assure (among other things) enough money for governmental needs as well as enough to enterprisers to attain high levels of economic activity.

Chapters 8, 9, 10, and 11 discuss the monetary and banking systems of this country and their role in the exchange of the commodities and services comprising the national output.

The Price System

The prices of consumer goods, producer goods, and services form a system of prices. By and large in this economy the production of an economic good is regulated by its selling price and its cost of production relative to other prices and costs. In turn its cost is composed of the prices of materials and services used in its production. Thus there exist changing relationships among many prices which regulate the production of types of things and their relative amounts.

The bulk of national output in this country is produced by private business in expectation of profit. A producer makes something out of his business venture only if the price at which he sells his output is high enough to cover his costs and leave something for him as profit. This puts the problem simply, but this is virtually all that counts in the calculations of the producer. Prices are the inducements, and costs are the obstacles, to production; if the producer expects these obstacles to be surmounted, he will produce and thereby contribute to national output, national income, and employment. What is produced, and how much of each thing is produced, are decisions determined by the relations of prices.

That is to say, the *composition of the national output* is mainly determined by the producers' expectations of production costs, sales prices, and volume of sales. As each producer engages in production in accordance with these profit expectations, economic resources are hired, bought, or employed and put to work. From these actions comes the *pattern of resource allocation,* which means the relative amounts of labor, capital, and other resources devoted to the production of each type of goods—wheat, books, lumber, greeting cards, aluminum, cigarettes, etc. The investigation of resource allocation is one of the central interests in the field of economics and for this reason the determination of relative prices is important; this latter subject is treated in Chapters 12 through 16.

The profit expectations of producers govern not only the composition of national output but also in large measure the level of national output, income, and employment. When profit expectations generally are favorable—i.e., when cost, price, and sales expectations of most producers are optimistic—total output rises; when these expectations are gloomy, total output is depressed. In a capitalist economy such as the United States and

Canada, both the composition and the level of national output are governed principally by producers' expectations of their costs, selling prices, volume of sales, and profit margins.

International Exchange

The exchange of commodities and services is world-wide. Goods produced in this country are carried to all parts of the world and a great many things produced elsewhere are imported. Foreign trade is an outgrowth of international specialization. There is, therefore, an international allocation of productive resources, achieved in part by price-cost relationships throughout the world and by governmental regulations controlling imports and exports. The exchange of commodities and services through world commerce is, like domestic exchange, carried out by means of money; in this case, however, the money used comprises American dollars, English pounds, Mexican pesos, and the monies of all other countries. International trade and finance are discussed in Chapters 34, 35, and 36 where they are related to the major problems of this book, namely, the composition of national output and the level of national output, national income, and employment.

THE DISTRIBUTION OF INCOME

The distribution of income, like production and exchange, is a function which is performed in any economic system. Every economy must work out methods of distributing income. Output and therefore income is limited; at any one time to give more to one person or group means to give less to others. How, then, should income be distributed? If more for one means less for another, on what basis shall this decision be made? All wants cannot be satisfied. Whose wants shall be more fully satisfied and whose less fully?

This is the problem of the distribution of *real* income. Real income has been defined as commodities and services; the distribution of these goods among the members of society is therefore the distribution of real income. The material well-being of persons and groups depends chiefly upon their relative shares of the national output. Those who have for themselves a good share of output are clearly better off in a material sense than those whose share is much less. It is for this reason that the most contentious problems of society arise from the distribution of income; throughout the ages the "haves" have fought to maintain their favored position and the

"have nots" to gain for themselves a larger share of real income and so to become "haves."

In this country the distribution of real income is governed largely by the distribution of *money* income—wages, salaries, profits, rent, etc. These incomes constitute claims to commodities and services, i.e., they constitute the power to purchase these things. This method of distribution involves the following principal steps: (1) that the money income is paid primarily to those who take part in the producing of national output; (2) that the amounts of income thus received are determined in general by each person's or group's importance in the production process; (3) that the total incomes paid out in the productive process approximate the total value (or amount) of national output: and therefore (4) that national output is largely distributed in accordance with the productive effort furnished by each person, either in the form of personal services or the services of the things he owns.

Money incomes depend on prices paid for the productive services of economic resources. A wage is the price paid for the services of a worker, rent is the price paid for the services (i.e., use) of land and buildings, etc. These prices are determined in a manner similar to the prices of commodities. Therefore as "prices of productive services," income distribution is analyzed as an aspect of price determination in Chapters 17 through 21.

The distribution of income among people is customarily termed the *personal distribution* of income. Attention is centered on these questions: How is the total distributed among income groups? What is the extent of income inequality? What are the principal causes of inequality and the principal consequences? What may be done to lessen the great spread between high and low income? Chapter 37 summarizes statistics on income distribution showing the numbers of income receivers in various income groups, ranging from the very low to the very high incomes; this may be termed the pattern of personal income distribution. The economic significance of the pattern of personal income distribution is analyzed in Chapter 38.

Personal income distribution in part governs the saving-spending activities of the nation's income receivers. Income statistics show that the bulk of the nation's personal savings is set aside by a relatively small fraction of income receivers—the majority spend all or virtually all of their incomes. It is probable that a lessening of income inequality would reduce savings and increase consumer spending, while an increase in inequality would have the opposite effect. As total national income increases, individual incomes in general rise, and savings increase relatively more than do con-

sumer expenditures. As total income declines, savings as a whole decline relatively more than consumer expenditures.

The way in which national income is distributed also influences the pattern of consumer expenditures, i.e., the proportionate expenditures for such things as food, rent, clothing, recreation, and education. Nearly all the income of the low-income groups is spent for food, rent, and clothing; as incomes increase, however, these items account for a declining portion of consumer expenditure while recreation, education, and medical outlays account for an increasing fraction. Changes in income distribution therefore tend to alter the pattern of consumer expenditure. Since consumer demand or expenditure plays a large part in guiding producers, the composition of national output and the allocation of resources are accordingly regulated in part by income distribution.

The personal distribution of income, therefore, is intimately related to the central theoretical problems of this book, namely, the causes and conditions regulating (1) the level of national output, income, and employment, and (2) the composition of the national output. In turn, the level of total income and changes in the level profoundly affect the distribution of income.

THE PROCESS OF CONSUMPTION

The fourth major process in the economy is the consumption of output. It has been seen that the composition of the national output covers a great range of commodities and services. Most of this output consists of consumer goods which are ready for the final stage in the economic process, i.e., consumption or the satisfaction of wants. Each component of national output is limited in relation to the wants for it. Therefore, the questions at hand are: How are consumer goods and services apportioned among the population? How is the consumption of scarce things decided? What determines how much of this or how much of that will be consumed? and who shall consume this or that?

In some way, each economic system must establish a method or methods of guiding and limiting consumption. The American economic system does so largely by means of prices, which are taken to include the incomes people receive. Things that are very scarce (in relation to the desire for them) sell at high prices, and only the consumers who can afford such things are able to satisfy their wants for them. Take, for example, platinum mink coats. Very few of these coats are made each year; in this case the wants are many but only few are destined to be satisfied. And whose wants? Those

of the few people who are able and willing to pay $30,000 or more for such a coat. The limited supply is effectively apportioned among consumers, for only a handful are interested in the article at this price. The consumption of so scarce an item could, of course, be regulated in other ways. A lottery could be established for the disposal of the coats. Or government could apportion them free or at a nominal price to the wives of top-ranking public officials and service officers, or perhaps present them to women who had given birth to some large number of children. Many other criteria could possibly serve as bases for the apportionment of such a scarce item.

On the other hand, many items in the national output are produced in large amounts. Bread is an example. Its low price assures a widespread consumption. Scarcely any family need go without some bread, and most families will consume about as much as they want of it. Other means of assuring a wide apportionment of bread could be adopted. For example, public authority could allot a certain amount to each person under a rationing system. In either case consumption is apportioned; in the one case, no more is consumed than people are able and willing to pay for, in the other, no more than the rationing authorities permit.

Whether it be platinum mink coats, bread, or any other item in the national output, the basic problem is the same in all economic systems; goods and services must be allocated or apportioned in some manner, for the amounts of these things are not adequate to satisfy all wants. That is, means are established to restrict total consumption to the available amounts, and by the same token the consumption of each person (or family) is restricted.

THE PRINCIPAL ECONOMIC DECISIONS

The processes described above—the *production* and *exchange* of national output, the *distribution* of national income, and the *consumption* or use of the commodities and services thus produced—all center about the making and disposing of money incomes. The processes depend on innumerable economic decisions made by producers, consumers, government, and organized groups in carrying on their roles in the economy. Economic behavior involves the selection of alternatives; not having a plethora of want-satisfying things, consumers must choose between this or that, or between more of this and less of that. Given their limited incomes people must choose between spending more and saving less, or spending less and saving more. In their turn, producers choose among the many things that

may be produced, and the possible amounts. In its economic functions, government too makes decisions: Should taxes be raised or lowered? Should government spend more or less and for what purposes? Should government action strengthen this group against that group?

The decisions to spend, save, tax, borrow, to employ resources and make things, are interrelated and form the pattern of economic behavior. In effect they produce a unity among the processes of producing and exchanging national output, distributing national income, and consuming national output.

In whose hands should rest the responsibility for making economic decisions? What division of responsibility should be made among government, employer, employees, and consumer? Should the basic decisions of the economy be made largely by government authorities, or largely by private persons or groups? In Soviet Russia these questions are answered in one way, in the United States in another way. In both countries, the same kinds of decisions must be made: commodities and services must be produced and the output must in some way be distributed so that some human wants may be satisfied.

As has already been indicated, economic decisions that mold the character and determine the activities of the economy may be classified as (1) decisions of producers, (2) decisions of income receivers, and (3) decisions of government. This classification is brief but adequate for present purposes.

Producers' decisions lead to the employment of economic resources for the production of commodities and services. These are decisions to spend money for hiring of labor, obtaining materials, procuring or hiring capital, and for insurance, public-utility services, and other miscellaneous items. Although it is arbitrary to establish a scale of importance among economic activities, there is much to be said for placing producers' decisions foremost. They play an initiating role in the economy; from these decisions there is initiated the stream of commodities and services making up the national output and also the stream of income payments making up the bulk of national income or purchasing power in the hands of the public.

In the American economy most of the decisions to produce are made by private enterprise in search of profit; corporations, individual enterprisers, and partnerships are forms of organization in which most producers in this country operate: they are described in Chapters 25, 26, and 27. The expectation of profitable sale or use prompts the decisions of these producers. This has been variously termed the *profit motive*, the *pecuniary*

motive, or the *self-interest motive,* and the system of production based thereon is commonly termed the *private enterprise system* or *capitalism.*

The role of producers' decisions in organizing and guiding the productive process may be sketched only briefly here. In deciding what kinds of consumer goods to produce and in what amounts producers appraise or gauge their markets, i.e., the diverse wants of the consuming public, the income in the hands of this public, and the willingness of people to spend money for the purchase of goods. Such appraisals result in given outputs of food, shoes, suits, handbags, radios, automobiles, cigarette lighters, and countless other things, i.e., in the composition of consumer goods in the national output. The allocation of a large part of the nation's resources occurs in the same way; producers' decisions determine the number of people employed in the various branches of industry. Thus the profit motive supplies the driving power, and the profit expectations the guideposts, for the productive mechanism in a capitalist economy.

Before a sizeable volume of consumer goods can be produced, however, producers must build plants, install machines and equipment, and acquire other forms of capital facilities. Decisions to spend money for these purposes are termed *investment decisions.* Casual observation will disclose the particular importance of such decisions in the American economy, for the American productive system is pre-eminently a capital-using system. Investment decisions are made on the bases of producers' appraisals of future consumer demands. Decisions to invest, together with decisions to produce consumer goods for current consumption, dominate the allocation of resources, the types of goods and services produced, and the total amount of national output produced.

From this discussion the role of consumers' (income receivers') decisions is apparent. Consumers exert their influence simply by buying what they want and can afford and rejecting all else. The market place has been likened to a great polling booth where consumers' dollars are the ballots. The many products campaign for the support of the consumers whose ballots consign some to that large and costly graveyard of unwanted products, the rest to a life of uncertain duration and of varying degrees of success. Freedom of consumer choice or decision is a central characteristic of capitalism; except for a limited list of legally forbidden things, consumers may spend as their incomes permit and their tastes dictate. Producers, of course, attempt to influence these tastes through advertising. The results of consumers' choices may not always satisfy some standards of morality, good taste, and nutrition, but the fact remains that consumer expenditure measures wants, taking consumers and their incomes as they

are. Producers, sensitive to the market, adjust production accordingly, for their decisions are basically derived from actual and expected consumer decisions.

Labor unions participate in the making of economic decisions by influencing wage rates, hours of work, and other conditions of their members. The growth, characteristics, and activities of unions are discussed in Chapters 28 and 29.

Governmental decisions, or governmental policy, enter into the economic scene in many ways. Of special importance are fiscal activities—the taxing and spending policies of government. In the years 1941-46 nearly one third of the national income was absorbed by federal, state, and local taxes. The Federal government purchased nearly half of the total national output in 1943-45, most of it for war purposes. Wartime expenditures exceeded tax revenue by about 50 billion dollars annually and the federal public debt increased accordingly. These fiscal developments arose out of the exigencies of war but their economic and financial effects will be felt indefinitely into the future. Both governmental expenditures and tax receipts in the years ahead are likely to be far above prewar levels; for this reason the economic aspects of government financial activities are now of especial importance. In Chapters 31, 32, and 33 the tax system, governmental expenditures, and public-debt problems are described; and their effects on income distribution, the composition of national output, and the level of output and employment are analyzed.

In still other ways governmental policy affects the use of economic resources, types of output, and the distribution of income. Minimum-wage legislation, old-age and unemployment insurance, the regulation of banks, and the issuance of corporation stocks and bonds, control over agricultural production and prices—all are examples of far-reaching governmental intervention in the economic processes.

Economic Decisions as an Expression of Economic Systems

The decisions that activate this economic system have been traditionally individualistic in origin. Recent times, however, have witnessed the growing power of organized groups that assume for themselves the function of making economic decisions. Labor-union officials speak for "labor," farm organization officials for "farmers," large corporations or associations of firms for the other business firms in their industries. Decisions relating to prices, production, wages, and investment tend to emanate more and more from organized groups or from the frequent conflicts between such

groups. Government in its fiscal and regulatory policies shifts many decisions from private to public hands. The traditional individualism in the making of economic decisions has therefore become modified—sometimes strengthened, sometimes supplanted, and at least occasionally perverted—by the action of organized pressure groups and government in the economic sphere.

Decisions to invest, employ, produce, spend, and save do not rest largely on individual or private group interests in other types of economic systems. The principal decisions in Soviet Russia are made by governmental authority. The contrast between the Soviet system and the private enterprise system of this country is brought out in Chapters 30, 41, and 42. It rests essentially on the location of the power to make economic decisions. In capitalism, power is essentially decentralized; each producer and consumer plays an individual part and decisions are prompted by motives private to those who make them. In communism, on the other hand, economic decisions are centralized in the governing bureaucracy and are prompted by motives common in the bureaucracy. However these decisions are made, the central economic problem is the same: to get the best results from limited economic resources for the satisfaction of the wants which the system permits to be expressed. In the American economy these wants are largely for the individual to express in accordance with his own ends or goals; in communism the wants that may be satisfied are largely for the governmental bureaucracy to establish and measure. In short, the want-satisfying decisions (which encompass directly or indirectly all economic decisions) are decentralized in one and centralized in the other type of economic system.

CHAPTER 2

Change and Growth in the Economy

Every economic system establishes some means of organizing and regulating the basic economic activities essential to the satisfaction of wants. By the *productive* process natural resources, capital, and human effort are

arranged so as to obtain the things that are deemed necessary or desirable by the society. This output must be *exchanged,* for in all forms of social life production is more or less specialized; hardly anyone produces all the things he needs and no more. The production of output creates income which must be *distributed.* The production and exchange of goods and services and the consequent distribution of income consummate in the use of output, which is primarily the *consumption* of goods and services. Because output is scarce and is insufficient to satisfy all wants, some method of regulating consumption is requisite to every economic system. These constitute, then, the basic activities which have occurred in various forms in the economic systems of both the past and present.

In this chapter there are traced the main outlines of these economic activities as they developed and changed in the economy of western Europe from the feudal period to the present. A brief review of economic history will aid in the understanding of modern economic organization in several ways. It is obvious that the economy of the United States emerged from the western European social, political, and economic environment and must be analyzed partly in terms of historical development. Moreover, reviewing economic history makes clear the fact of change in economic life. From the Middle Ages to the present, western Europe went through a series of revolutionary economic developments; wars, political reformations and revolutions, technological innovations, movements toward geographical expansion, population increases, ideological conflicts—all have transformed the ways in which people make a living. Economic systems are never static; the channels of development and the directions of change in the past not only illuminate the nature of modern economies but also provide a basis for regarding the prospects of the future. Finally, a long period of growth and expansion in output and income is perhaps the most significant feature of the economy of the West, and particularly of the United States; again, it is necessary to examine the historical process in order to encompass a full understanding of this fact.

THE ECONOMY OF THE MIDDLE AGES

The economic world of western Europe five or six hundred years ago bears little resemblance to that prevailing today. Indeed, if a medieval man were to visit present-day France or England he would find but a few vestiges—surviving chiefly for ceremonial purposes—of the economy in which he lived and worked. The predominantly agricultural feudal economy, with only a limited handicraft industry, has been replaced by a

system with world-wide economic relationships in which are found great mechanized industries and large trading and financial centers. The local economic self-sufficiency of feudalism has vanished. A medieval man might now find his isolated village a town or city, dependent upon all Europe and much of the world for many of the things its inhabitants consume. Instead of the local exchange of necessities and meager comforts through barter, he would find a complex marketing and monetary system. The slow tempo of medieval life, rigidly regulated by religion, law, and custom, has been transformed into the modern dynamic of speed and progress. Most striking of all changes greeting this medieval visitor would undoubtedly be the great plenitude prevailing in ordinary times even among the poor. And America, unknown in medieval times, would afford him still greater contrasts than western Europe in opulence, in the nature of economic organization, and in the dynamic character of economic life.

No more than a brief survey of the economic elements of the medieval period can be attempted here. A cursory glance can bring forth only the dominant characteristics of so long a period as that ranging from the sixth or seventh century to the fifteenth century, and so diversified as to embrace both a well-developed commercialism in the Mediterranean area and a predominantly agricultural economy in Continental Europe.

The Conception of Economic Activity

To the medieval mind economic activity was an essential but subordinate aspect of human behavior. The main purpose in life was not to "get ahead" in a financial or worldly sense. The prevailing philosophy of the period and the attitudes and beliefs of people in general were dominated by the theology and institutional power of the Church. Worldly possessions and interests were subordinated by this theology to the main purpose in life, which was to attain salvation. To be sure, the economic tasks of each community had to be performed: necessary goods and services were produced and to a limited extent exchanged, income was distributed, and economic goods were consumed. But these economic activities were restricted. The observance of scores of holy days in the course of the year curtailed production; fulfilling spiritual needs was held superior to furthering material satisfaction by increased production. Moreover, although the exchange of goods was of course essential, the prices of goods in exchange were ruled by considerations of justice and fairness and not material gain. The individual who sought to increase his income and accumulate wealth roused suspicion. The rich man was considered to have gained his wealth

at the expense of others and thereby to have lessened his chances of salvation. As for consumption, the prevailing view cautioned moderation; the pleasures of the flesh diverted attention from man's only meaningful earthly purpose—to seek grace and prepare himself for salvation.

However, there were flagrant violations of the spirit and letter of the moral code. The arts by which men grow rich were well-known and often practiced, but still received no social approbation. In discussing the prevalence of avarice and greed in high places, Tawney observes that medieval men "called these vices by their right names, and had not learned to persuade themselves that greed was enterprise and avarice economy." [1] In feudal times, unlike a later period, economic behavior was judged on moral grounds and not measured in terms of its material dividend.

The Structure of Wants

The organization of the productive process of an economy is guided in large measure by the wants it serves. In the medieval economy the satisfaction of wants, or consumption, was regulated by the laws and customs defining each person's place in society. This society was a hierarchical system of classes—a pyramid with a broad base made up of slaves, serfs, and artisans, a higher level composed of merchants and master craftsmen in the towns and feudal retainers in the country, and at the apex, the nobility of both the laity and the Church. Within each class further differences existed; the lower order of people revealed gradations as did the nobility. Each group had its assigned functions, hallowed by tradition and religion and enforced by law. The lower orders worked to provide society with its material requirements; the lay nobility and retainers looked after its defense and the administration of civil government. The clergy ministered to the spiritual needs and because of the enormous authority of the Church also participated in political administration. Each class, then, had well-defined duties, responsibilities, and privileges. Except for the clergy, which drew upon all classes to an extent, one's position in life and therefore his occupation depended principally upon the class into which he was born. Little chance prevailed to "better oneself"; the accident of birth determined one's place in the social system.

This social hierarchy entailed an ordering of human wants; want satisfaction, or consumption, was fixed chiefly by one's "station in life." The whole hierarchy—serfs, tenants, artisans, the minor nobility, the great nobility, and all the other functionaries in this ordered system—enjoyed estab-

[1] R. H. Tawney, *Religion and the Rise of Capitalism*, Harcourt, Brace, 1926, p. 61.

lished rights and performed their customary duties. Among the rights—but probably not the most important—was that of receiving a "living" in accordance with one's station. The clothes a man wore, the amount of food he consumed, the dwelling he occupied, the fancies he could indulge and the luxuries he enjoyed, were determined by the customs and law regulating his station in life. The structure of wants was therefore governed by political authority and heredity under the sanction of tradition; "free choice" played little or no part in the individual's actions. Thus controlled, the consumption of the bulk of the population was limited to the barest essentials, while those whom birth had favored fared well by the standards of the time. It might be said, of course, that the limited productive capacity could provide no more than custom decreed for consumption. The same influences, nevertheless, which restricted consumption also fettered the productive activity of the feudal period.

The Organization of Production

Productive activity throughout this period was rooted largely in agriculture.[2] Land was the great source of wealth. It was let out by the nobility under a complex system of tenure to the serfs and tenants who paid feudal dues for the privilege of working it and bound themselves to render service to the nobility. The villa or manor was the principal economic unit in rural areas, around which agricultural output and handicraft activities were organized. The bulk of the laboring population was thus tied to agriculture by feudal law; they worked their own small tracts only part of the time, for some of their time and effort they owed to the lord of the manor. The crops produced, the methods of tillage and harvesting, and the other tasks performed on the lord's domain were rigidly controlled by custom. Indeed, little else was known except the backward farming practices handed down from father to son. Crafts were conducted similarly. Productive methods were static and, judged by modern standards, primitive. Most of the power used was human, aided by scrawny work animals; "labor-saving" devices were of little importance. Innovation in farming and in the handicrafts—if it occurred to many at all—encountered the opposition of entrenched ideas and practices. Thus were production and consumption alike organized and regulated by the same kind of rules.

Trade and handicraft industries in the towns were restricted by custom

[2] In the Mediterranean an extensive commercial trade flourished after the ninth century, but its area was restricted and it did not substantially affect the continental economic structure.

and also by the power of the *guilds*. Guilds were local organizations of merchants (merchant guilds) and craftsmen (craft guilds). Their purpose was partly to assure to their members a monopoly of the trade and industry in their respective towns by keeping "foreigners" out entirely or at least by closely restricting their activities. Guilds also regulated the activities of their members, seeing that each received his "fair share" of the local business and no more. Craft guilds determined the number of apprentices and journeymen each master craftsman could employ, the quantity and quality of the goods he produced, the price charged for the product, and the amount of raw materials he could buy. The craft guild performed important services. It was the only means of training workers for skilled occupations; and its apprenticeship regulations also assured high quality of goods produced.

The regulation of productive activity imposed by craft and merchant guilds was reinforced by the tenets and methods of the Church. Its doctrines insisted upon fairness and justice in economic activity. The system of economic ethics it advanced, concurrent with prevailing custom and public sentiment, created an atmosphere that discouraged the expansion of business interests during the period. Trade was tolerated uneasily as a necessary part of the life of the community; industry was regarded as necessary to satisfy the customary needs of the community. No more was granted. The subsequent growth of trade and industry found little or no ecclesiastical support. Support instead came from secular authorities, and in the later medieval period particularly from the national governments which endeavored to build up political and military power by the acquisition of wealth.

The medieval productive process, both on the manor and in the towns, was bound by tradition, by local monopolistic influences of guilds, by the ethics of the Church. The spirit of the age was not economic; while the age was politically turbulent it was economically quiescent. Production and trade sufficient to maintain the traditional hierarchical standards of income and consumption was enough. Whether with the resources at hand medieval society could have attained higher levels of production, income, and consumption is a matter of conjecture.[3] It is significant that there seemed to be little interest in the question at the time.

[3] Max Weber in his provocative book *The Protestant Ethic and the Spirit of Capitalism* (London, 1930) argues that it could not; the Protestant Reformation alone supplied the spiritual drive behind the subsequent growth of capitalism. R. H. Tawney in his *Religion and the Rise of Capitalism* (Harcourt, Brace, 1926) holds that the dissolution of feudalism and the rise of capitalism had no religious cause but both developments were stimulated by the Reformation. Students interested in this large question are urged to consult these important works.

The Exchange Process

Like other elements in the economic system the exchange of goods and services in the medieval period was hedged about by moral considerations; and it was limited further by the paucity of money and the difficulties of transporting goods in trade as well as by the locally self-sufficient character of the economy. Purchase and sale were in large measure subject to the Church doctrine of *Just Price*, which established the standard price considered just when buyer and seller exchanged equivalents. The price of a product was expected to equal roughly the cost of production; costs included a "fair wage" and a "fair profit." It was agreed that a person was entitled to a fair wage and a fair profit, and therefore to a fair or just price for his goods. In this way the consumer, too, was entitled to protection against unfair prices.

This concept subordinated the economic function of price as it is regarded in modern economics. Today prices are expected to equate the supply of and the demand for goods and services, and to effect an allocation of economic resources and a composition of output appropriate to the demands of the market. The medieval view of price, however, did not wholly overlook this definition of its function. Of necessity supply-and-demand influences exerted a control over market price; but regulatory measures, both by authorities and accepted moral standards, tempered these influences. A seller was not to charge so high a price as he could get for his product, nor was he expected to take the least price a buyer was willing to pay.

While fairness in exchange was the ruling principle, many prices escaped the moral injunction. The purchase of foreign wares was not easily governed by rules of morality. What was the cost of producing silk or jewels brought in from the East? Obviously no measure of fairness was possible for such goods. Moreover, since only the rich bought silk and jewels there was no compelling reason to protect them against unscrupulous traders. As the scope of foreign trade grew, particularly in the Mediterranean area, prices more and more eluded the mandate of morality and came under the market forces of supply and demand. Throughout the feudal period, however, local producers and local consumers were accorded the protection of the Just Price concept as enforced by the Church, the guilds, and local authority. As the feudal economy disintegrated, exchange developed rapidly and prices came to be more freely determined by the forces of supply and demand.

The Distribution of Income

The division of income in feudal society was determined largely by a person's position in the social hierarchy. One of the class rights of a person was the right to an income sufficient to maintain him in his "station in life." Thus, a cobbler or a tailor was entitled to a certain income; he was entitled to a just price for his work, i.e., a price that gave him sufficient income to maintain him in his customary station in life. A just price in this definition implied a fair income—a fair wage or profit. The peasant working on the land had a right to a certain share of the produce, enough to maintain him in his station. Similarly, the lord of the manor had a right to a portion of the produce of the land as well as to dues from his retainers. Income distribution bore little or no relationship to the amount of work one did, or to whether one worked at all. Like consumption, it was regulated rigidly by one's position in the prevailing social and political hierarchy.

MERCANTILISM

At the time of the dissolution of the feudal economy and the emergence of the modern age, mercantilism began to dominate economic thinking. No well-defined limits can be given to this period. By the end of the seventeenth century mercantilism was fully developed, and a century later it was clearly declining. During this period economic policies developed hand in hand with the growth of national states. In its political aspect, mercantilism characterized the process of *state-building* and the quest for national power which arose from the political localism and military insecurity of the medieval period. In its economic aspect, mercantilism sought to encourage the *growth of capital,* and to create industry and trade which would provide the national state with the economic means of attaining military and political power. In effect the activities of the mercantile class were served by the architects of the national state. A profitable and far-flung trade benefited the separate interests of both the merchants and the rulers of the state.

Purely economic activities grew in importance and in respectability from the fifteenth century. Trade was foremost among these. Despite the suspicion cast upon it by the churchmen of the Middle Ages, the growth of trade was irresistible first in the Italian city states, then in the Low Countries and the Hanseatic cities, and finally, in all of western Europe. The

so-called Commercial Revolution of the sixteenth and seventeenth centuries marked the shift in trading supremacy from Italy to the Atlantic, and eventually to Great Britain. Mercantilist policy fostered and controlled this rise of commercialism; indeed, the strength of a state was measured largely by its merchant fleets and the gold and silver derived from trading.

The central economic doctrine and policy of mercantilism—the celebrated favorable balance of trade—bears witness to the exalted position accorded to trading and to the gold and silver specie derived from it. A *favorable balance of trade for a country meant that its exports were of a greater value than its imports, with the excess of exports over imports being received by the country in the form of gold and silver.* In this way the state provided itself with the financial basis of power and with the money required for the expansion of industry. Mercantilist governments therefore fostered export trade. This was done by granting monopoly privileges to trading companies, such as the particularly successful East India Company and the trading companies formed for the development of the American colonies. Besides political support, governments also provided direct financial aid to many trading companies. Export trade was often encouraged by means of bounties on exports paid from the national treasuries. Domestic industry was stimulated to provide the goods supplying the export markets. At the same time imports of goods, particularly manufactured goods, were discouraged and sometimes prohibited, as a measure to protect domestic industry and to assure an excess of exports over imports and therefore an inflow of gold and silver.

The growth of trade was affected, furthermore, by an increase in population throughout Europe and the establishment of towns and cities. Trade of necessity expanded as urban centers became less self-sufficient, and this necessity became a virtue in the philosophy of the mercantilists. A numerous population was of utmost importance in the view of the mercantilists, for it provided a cheap and abundant labor force for the growing industry as well as recruits for the armies of the national states.

An active trade inevitably entails a rise in production. By the seventeenth century mercantilist attention shifted more and more from trade measures to manufacturing methods. French mercantilism, for example, was largely concerned with the control and development of manufacturing. The English burdened industry with fewer regulations than the French, but nonetheless industrial growth, particularly in the textile industry, was a matter of national interest and some state control. In this period important changes had occurred in the organization of the manufacturing industries. The restrictive influences of the craft guilds had gradually weakened. Manufac-

turing, especially of new products and processes, increasingly shifted away from the towns to the country where much of it was carried on in the home under the "cottage system" or the "household system" of production. At this time, too, the factory system made its appearance on a modest scale, destined later with the use of water and steam power to rise to dominating importance.

Both trade and manufacturing, indeed all economic activities in the sixteenth and seventeenth centuries, felt the enormous impact of gold and silver from the New World. In Spain and Portugal the influx of gold and silver caused an inflationary rise of prices; the supply of money increased, but supplies of goods grew less rapidly or remained stable. As the specie flowed into other countries—by force of war, piracy, or favorable balances of trade—the inflationary influences spread. The effect was disruptive, as great price increases always are. But as a result great profits were made and manufacturing and commercial activity throughout much of western Europe was subsequently expanded. The old feudal agricultural economy, based largely on local needs and the barter system, collapsed. A "money economy" spread widely, i.e., trade based on the exchange of goods for money became prevalent. The local handicraft industries found wider markets as more money came into circulation and the number of trading companies grew. Thus it may be said that the mines of the New World financed the dissolution of the old feudal economy.

Economic life underwent revolutionary changes between the fifteenth and seventeenth centuries concomitantly with the political reorganizations of western Europe. The changes in politics and economic life were accompanied and to some degree caused by the religious upheaval of the Reformation. Moral restraints on trade and profit-making were weakened; commercialism became respectable and, at least to the Puritan's view, a manifestation of moral virtue. The spirit of the times had been transformed. Political power and rivalry, international competition in commerce and industry, the quest for additional and greater markets, for gold and silver, for economic expansion—these considerations dominated the rising middle class and the political groups who supported this class to assure the power of the national state.

Organization of the Economy

How was the new economic order of mercantilism organized and controlled? How were the central processes of production, exchange, distribution, and consumption regulated? The decline of feudalism and its eco-

nomic controls necessitated the development of other controls, for society perforce must find some means of regulating economic processes.

This regulation was in part based on the mercantilist conception relating economic activity to the interests of the state. There were, of course, differences of opinion as to what constituted the interests of the state. In the seventeenth century, there was no dearth of special pleaders who interpreted the "interest of the nation" in their own lights, just as organized groups today claim to speak for the interests of "the people" or "the social welfare" or "humanity," each interpreting these abstractions to suit its own purposes. The aim of mercantilism, however, was limited in scope, and the means of attaining it were deemed to be relatively clear-cut, namely by strengthening the economic position of one's own state and weakening the economic position of all potential rivals. In this way the political power of the individual state was enhanced. Obviously, the concepts of "world prosperity" and international economic collaboration were foreign to the thinking of the period. Indeed, to the mercantilist, world prosperity could mean only that there were many strong economic and military adversaries, that war was inevitable, and alliances nothing more than temporary military expediencies.

The Productive Process

By such reasoning the importance of the productive process and the wealth it created was measured in relation to the aims of the state. The desirable productive system was one that produced goods and services useful to the strength of the sovereign; the welfare of the individual per se was of little consequence. Those in power protected their own individual interests while ministering to that of their prince. As for the others, the "lower orders of mankind," it was enough that they subsisted to work, to propagate their kind, and when necessary to fight—all for the state's interests. The productive process was not directed for their material well-being.

The desirable composition of national output was one which provided: (1) a maximum of manufactured goods for export; (2) the necessary goods and equipment for military and naval purposes; (3) the subsistence of the population. Mercantilist governments, however, did not trust private enterprise to produce all that was needed for the national interests. Intervention was therefore the order of the day. The manufacture of goods necessary for military and naval purposes or desirable for export purposes was frequently subsidized by royal treasuries, and the price and quality of such goods were commonly subject to state regulation. Governments,

too, resorted to various pressures to get the production of things deemed essential to national purposes. Frederick the Great coerced his subjects to engage in productive activity essential to the strength of his kingdom; and he established state industries to produce military supplies and equipment because private producers were unable or unwilling to engage in this work. Both France and Britain encouraged the iron-making industry as well as ordnance. The French government introduced and encouraged the production of silk, the English the production of woolen textiles for domestic use and for export purposes. The production of tools and machines for manufacturing was encouraged by all mercantilist governments. National states employed various means to stimulate production and encourage the founding of new industries. Governmental subsidies, governmental establishment and ownership of industry, favorable price controls, export bounties on manufactured goods, prohibition or limitation of imported manufactured goods to protect domestic industry, the encouragement of invention, and threats and coercion were all used. Governments, too, resorted to the theft of machines from abroad, the enticement of foreign skilled workers, and the abduction of such foreign workers, when inducements failed, in order to build up new manufacturing processes and industries. At that time the skilled artisans of defeated foreign countries enjoyed a position akin to that of the scientists of defeated nations in modern war. The simple machines of that day, like the industrial equipment, secrets, and patents of the modern world, were prized as means of building up the productive power of the victor in war.

In all such activities governments supplemented private enterprise as the means of producing the desirable types and amounts of goods. In other directions, private enterprise was often restricted when its activities were considered harmful (e.g., the export of arms) or merely unnecessary (e.g., the production of luxuries). Thus the structure or character of the productive process, and therefore the allocation of economic resources and the composition of national output, became matters of governmental "responsibility." The means of attaining the desirable productive results were partly price-cost relationships, and partly such interventionist measures as have been described. In the course of time, as the power of mercantilist thought and policy weakened, profits became the principal guide to productive activity and governmental regulation declined in importance.

Not only did mercantilist policy govern the types of goods produced but also the total amount of production. Many of the governmental measures already described directly stimulated increased amounts of investment, employment, and national output. These measures, together with

the widening of overseas and domestic markets, stimulated private investment and the expansion of output. Although many of the mercantilist regulations and policies undoubtedly acted as deterrents to investment and production, their origin in general was prompted to foster the growth of the economy.

Trade, as an aspect of the productive process, was similarly regulated and guided by mercantilist policy. The freedom of internal trade was generally espoused as a means of strengthening the economy, but foreign trade was subject to regulation. In the first place, exports were stimulated and imports restricted, in order to obtain a favorable balance of trade and the influx of gold and silver. For this reason imports of raw materials for domestic industry were encouraged and their exportation discouraged, while imports of manufactured goods were discouraged and their exportation encouraged. Governments forbade or restricted the export of munitions, and encouraged or sometimes required shipmasters to import such items. Manufacturing tools and machines were placed in a similar category. In these trade controls the mercantilists did not place faith in private interest; "free trade" in the literal sense was not trusted. "Trade indeed will find its own channel, but it may be to the ruin of the Nation, if not Regulated," expresses the typical mercantilist view.[4] The mercantilist writer Mun distinguished between the commercial gains of the country, of the merchant, and of the king; any one of these might be at the expense of the others.

The Process of Exchange

The growth of mercantilist influence resulted in the dissolution of the medieval concept of the exchange process. The Just Price doctrine, it has been shown, conceived this process to be one in which justice and morality should govern; the control of price was an aspect of the administration of justice. The extension of the division of labor and a growth in markets, especially through foreign trade, weakened the moral aspects of exchange at the time feudalism gave way to the commercial era. This is not to say that mercantilism allowed "free exchange" to occur, i.e., the unfettered buying and selling of goods at prices determined by influences of supply and demand. The mercantilists regulated many prices and other conditions of exchange in accordance with their economic aims. Nevertheless the "price system" in its modern sense developed in this period, and mer-

[4] Sir Francis Brewster, *New Essays on Trade*, 1702. Quoted from Heckscher, *Mercantilism*, George Allen and Unwin (London, 1935).

cantilist literature undertook an analysis of the supply-and-demand influences regulating price. The celebrated "law of supply and demand" took form in this period. The price of a product was determined by the amount offered for sale and the demand for it. As these changed, prices changed; if supply were large or demand weak, price would fall; if supply were short or demand strong, price would rise. These relationships were regarded as a "natural law" and some of the later mercantilists began to question the ability and the desirability of governmental regulation of matters subject to the jurisdiction of nature.

Income Distribution

Similarly, the growth of the division of labor, of commercialism in trade, manufacturing, and agriculture, of urban life, and of the use of money broke down many of the feudal arrangements that governed the distribution of income in the earlier period. Under mercantilism money incomes became common; the economic activity of the individual was directed largely toward getting a money income. In agriculture the rental contract had replaced the feudal dues and services, particularly in England. In manufacturing, trade, and agriculture the modern wage-earning class was clearly emerging. The merchant group was well established and was earning substantial profits in its trading activities. The manufacturing group, although less prominent than the merchants, was evolving out of the master-craftsman category. The "wages of the laboring poor," the "profits of stock or trade," and the rent of the landlord became a familiar division of money income in the seventeenth century. Increasingly the productive services of labor, capital, and land were bought and sold in their respective markets—the modern era of money income and money payment was forcing out the early system in which income was largely in the form of goods and services and rigidly controlled by custom, law, and heredity.

Mercantilist governments intervened extensively in the distributive process and thus affected the income pattern. The granting of trading monopolies to favored groups enabled them to gain large profits in foreign trade at the expense of those excluded. Manufacturing monopolies and special privileges similarly enabled the recipients of government favors to reap large incomes. Intervention, stimulation, and restriction on the part of government in the productive process created income advantages for some, disadvantages for others. The widespread control of wages by public authority led to the fixing of these incomes at low levels—the so-called subsistence wage idea, i.e., payment to wage earners, based on the mini-

mum needs, found general support among mercantilists. Despite these controls, however, the character of the economic processes in this period led increasingly to a determination of income distribution—wages, profits, interest, and rent—by market influences. More and more, wage rates were governed by the number of wage earners in a locality and the demand of employers for labor, interest rates by the supply of money and the demand for its use, rent by the demand for the use of land; in short, the factors of production had their markets. The rates of remuneration paid to the factors were governed in large measure by supply and demand as governmental control gradually weakened in the late mercantilist period.

The Process of Consumption

As noted above, mercantilism subordinated consumption—the satisfaction of individual wants—to other purposes. Increased consumption in the home country meant fewer goods available for export. The spending of money for consumption meant fewer funds available to "set productive labor to work." Consumption by the public was therefore subjected to a number of direct and indirect controls, the most effective restriction of which was low incomes. As has been stated, wages generally were regulated by public authority at the subsistence level; [5] wages above this level would encourage high living and a wanton waste of money on luxuries. Keeping wages at the subsistence level had other advantages as well. It encouraged industry; if 16 hours a day were required to earn a subsistence wage, workers would have no choice but to work that long. If they earned a subsistence wage in 10 hours, they would work only 10 hours, thus limiting national output. Hunger was held to be a "spur to industry"; low wages encouraged production and discouraged consumption, thus increasing the "surplus" for export. Furthermore, low wages meant low costs of production, and low costs enabled a country to compete advantageously in the sale of its goods in the foreign market.

Consumption was controlled in other ways. Sumptuary legislation was common, although it was poorly enforced. These laws forbade expenditures on certain types of goods, usually those deemed luxuries and frivolous in character. Such legislation was, of course, prevalent in the medieval period as a means of keeping the lower orders in their proper places in

[5] The theory that tied wages to subsistence is kin to a modern version which relates them to "cost of living." While both share the same theoretical limitations, they differ greatly sociologically. The former sought to keep wages low, but the cost-of-living version is used by its proponents as an argument to increase wages as living costs rise but not to lower wages as living costs fall.

the social (consumption) scale. Mercantilism intensified this legislation, partly to keep social classes fixed but mainly for economic reasons, i.e., to restrain consumer spending on luxury goods. Occasionally, however, regulation attempted to force consumption of particular commodities—produced in monopoly establishments owned by the state itself or by court favorites. At the same time, the mercantilists sought to keep foreign consumer goods out of the country. As a consequence the satisfaction of consumer wants was restrained and regulated, especially that of the lower orders.

Controls over consumption provided a frame of reference for the allocation of economic resources. Some industries were "more useful," others "less useful," and still others detrimental to state interests. Industries that catered to trivial wants were in the last category; sumptuary laws and subsistence wages were designed to stifle demand and throttle such industries. For example, the rich were exhorted to save, to eschew luxury spending, and to confine their consumption to domestically produced goods.

Mercantilism, like the feudal economy, established its own means of guiding and regulating the processes of production, distribution, and consumption. In the sixteenth and seventeenth centuries, this regulation of the economic processes was *politically* undertaken. A resemblance between mercantilist control over the central economic processes and the wartime regulation of the economy in the United States becomes apparent. Here, too, the productive system was made to serve a national aim: to win the war. In accordance with this aim, goods and services were classified as "essential" and "nonessential," with numerous subtypes. The essential industries were given priorities to get labor and materials; governmental funds were used to expand industry. Nonessential industries were given no help, and in some cases their production was prohibited entirely. Consumption was regulated in accordance with the plan of production; many goods were rationed, incomes were subjected to some, if minor, controls, taxes and governmental borrowing took a large share of the incomes. In principle nothing new was added during the recent war to the mercantilist government practices; but in operation wartime economic controls were more comprehensive, more detailed, more single-minded, and much more efficient than the measures of mercantilism.

ECONOMIC LIBERALISM

Mercantilist ideas and policies gave way in the course of time. The evidence of their decline was abundant in Britain during the eighteenth century. France abandoned the restrictive system more reluctantly; the

revolution administered the *coup de grâce,* although the forces of change were stirring long before 1789. In Germany and other parts of central Europe, the policy of governmental economic guidance was never wholly abandoned, but only weakened for a while during the nineteenth century. In America, mercantilist policy and ideas were present at its colonial birth; indeed, the enforcement of these restrictive policies was one of the causes of the colonists' demand for independence. Partly for this reason, and because this was a new country, the traditions of British and European mercantilism never took firm root in this soil.

The decline of mercantilism ushered in the ideas and policy of *economic liberalism.* Under this policy the individual was accorded the freedom to buy and sell, to produce and consume, as his fancy and finances permitted. Individualism, motivated by a rational self-interest, became the keynote in economic activity. Government receded in importance; *laissez faire,* or "let alone," became the watchword.[6] Let trade be free to move as it will; remove governmental restrictions on production, employment, prices, wages, consumption; abandon governmental privileges to trade and production monopolies. The goal was the establishment of a "free economy" in which private persons—consumers, wage earners, enterprisers, the owners of land and funds—would make their economic decisions on the basis of their own interests, guided by market-price relationships. Economic liberalism conceived an economy in which all persons would be free to dispose their incomes, their services, and their property without the compulsion or restriction of authority—governmental, feudal, or religious. In place of the authoritarian control of economic decisions by law, regulation, and tradition, economic decisions came to be controlled by *economic* relationships.

As with previous great social changes, economic liberalism grew slowly and unevenly throughout western Europe. The commercial and manufacturing interests of the seventeenth century, and particularly during the eighteenth century in Britain, became increasingly restive under the regulatory hand of government. In Britain, and to a lesser extent elsewhere, these groups had risen to a position of social, financial, and political power. Increasingly merchants protested against regulations and the granting of privileges to a favored few. Sentiment toward "free trade" was cogently expressed by some leading mercantilists in the late seventeenth century. By the eighteenth century violations of mercantilist law were widespread and publicly condoned. A growing segment of the business community saw

[6] The phrase *laissez faire, laissez passer* was used by the physiocrats in France during the middle of the eighteenth century, when as champions of economic freedom they demanded government to "let people do, make, and go" as people wished; in short, "let them alone."

in economic freedom a greater scope for profitable operations. Business opportunities appeared bountiful; economic expansion domestically and overseas was certain. Where governmental restrictions limited these opportunities, clamor arose for their removal. Manufacturers resisted regulations that hampered their productive activity. The growth of trade created an increasing demand for their products and they sought to satisfy this demand in their own way and in terms of their own interests. Wage earners frequently manifested their opposition to mercantile and feudal restrictions upon their freedom of movement and their wages by migration, disobedience, and violence. Liberalism became a social movement drawing support from many segments of society.

These practical interests found support in the philosophy of the eighteenth century which posited a system of *natural laws* governing social behavior. Natural law was held superior to man-made law: it was the duty of government to conform to the dictates of natural law. In the third quarter of the eighteenth century the physiocrats in France proclaimed the ascendancy of the natural order and natural law. With somewhat less didactical conviction Adam Smith urged conformity with the "obvious and simple system of natural liberty." Economic behavior was believed to be natural and spontaneous; it would be guided to the public benefit by the self-interest of man, provided the obstructive tangle of feudal vestiges and mercantilist institutions was cut away.

This doctrine was based on faith in a *natural harmony* inherent in economic phenomena and processes. The harmony existed between individual interests and social interests. By pursuing his own economic interest man was led (by what Adam Smith called the invisible hand) to further the public weal, even though this was no part of his intention. "The natural effort of every individual to better his own condition, when suffered to exert itself with freedom and security, is so powerful a principle, that it is alone, and without any assistance, not only capable of carrying on the society to wealth and prosperity, but of surmounting a hundred impertinent obstructions with which the folly of human laws too often encumbers its operations. . . ." [7] It was with these concepts of natural harmony and public welfare that economic liberalism departed fundamentally from mercantilism. The mercantilists identified public welfare with state power and saw no harmony of interests in the economic system. Only "the dextrous management of the skillful politician" guided private interests to further the public, i.e., state, aims.

[7] Adam Smith, *Wealth of Nations*, Cannan edition, p. 508.

In the *laissez-faire* view the freedom of the individual to pursue his own ends and adapt freely the means (resources) at his command to attain these ends spontaneously and naturally promoted the public welfare. The sum total of individual interests constituted the public interest. Public welfare, however, demands a well-regulated process of production, exchange, distribution, and consumption. How would this be assured if the time-honored regulations and controls were swept away? The answer to this question was supplied by the economic theorists; indeed the task of answering this question led to the study and development of systematic economic theory.

Organization of the Economy

Basic to the organization of the *laissez-faire* economy is the freedom of the individual to satisfy his wants as he chooses; i.e., the individual is accorded the right to follow whatever ends, aims, or purposes he desires. The hierarchy of wants existing for all people is for the individual to construct for himself, without intervention from the state, church, or feudal obligations and traditions. Specifically, he may spend his income for whatever things he wishes without restraint or compulsion. In spending his income his choices depend upon the importance he attaches to the things available to him for purchase, and upon their prices. The *supremacy of consumption* is thus asserted in the economic sphere by the *laissez-faire* doctrine.

Guided by this individual freedom of choice, total consumer expenditures flow into the markets for various types of goods and services. The prices consumers are willing to pay and the quantities of the various things they are willing and able to buy are guides to the productive process—to the allocation of labor, capital, and natural resources for the production of national output. Prices in the market and quantities purchased thus regulate the decisions of producers; things that can be sold at profitable prices are produced; things that cannot be sold profitably will not continue in production. When too little of something is produced in relation to consumer demand, consumers bid up the price, and the resulting large profit encourages producers to increase production. When too much is produced the price is depressed; producers suffer losses and output is reduced. As Adam Smith stated the matter, "the quantity of every commodity brought to market naturally suits itself to the effectual demand," i.e., to consumer preferences expressed by their ability and willingness to buy. In this sense, *the possessors of purchasing power dictate the kinds and amounts of things produced in a* laissez-faire *economy.* If producers are to stay in business they have no choice but to heed the market. In their turn, producers are

free to enter any field of productive activity that promises the prospect of profit, and likewise to withdraw from or shun any field not holding this promise. Similarly, wage earners and the owners of capital and land are free to offer their services and property to enterprisers offering them the best rewards. By this means all economic resources are mobile and may shift about among the various fields of productive activity in accordance with the income-spending decisions of consumers.

The productive processes during the late seventeenth and especially during the eighteenth century became increasingly specialized; the division of labor in different types of production and within a single industry or trade was extended under the stimulus of improved technology and an expanding market for goods. As already observed, division of labor necessitates the exchange of specialized products throughout the economy. During the several centuries preceding the ascendancy of economic liberalism, the exchange of goods and services had become primarily a matter of buying and selling, with money as the medium of exchange and price as the rate of exchange. In this period, however, mercantilist policies had encumbered exchange with many restrictions. Economic liberalism virtually freed the exchange process from existing restraints. Free markets and free trade came into dominance. In these markets consumer preferences were registered by the quantities of goods purchased at given prices, which in turn regulated the decisions of producers. A free exchange process thus became an integral, indeed a necessary, element of the liberal economy.

In the process of producing and exchanging goods the distribution of income takes place. Unlike the earlier periods in which incomes were controlled to varying extents by custom, social status, and law, economic liberalism gave control over income distribution to the market forces. That is to say, each person was free to obtain whatever price (income) he could get for his services or property. Wage earners were not by law limited to any given wage; they received whatever they could induce employers to pay for their services. The owners of land and money were legally free to use their holdings to earn an income, or to rent or lend them out to others at whatever rent or interest was obtainable. Enterprisers were free to go into any field of production they wished and to earn whatever profit their industry, ingenuity, foresight, and good luck afforded. In short, the productive services of labor, capital, and natural resources were bought and sold in their respective markets; the prices of these productive services were the wages, interest, and rent or royalties paid to wage earners and the owners of capital and land. The enterpriser who bought these services, and combined them into a productive enterprise, sold the products thus produced;

if he judged the market for these products accurately he received a profit as his reward.

The system of economic liberalism thus based the processes of production, exchange, distribution, and consumption on the free choices of individuals. Far from being chaotic the system developed its own means of regulation; prices of consumer goods limited the consumption thereof; prices of goods and services—in relation to their costs of production—regulated the production of each component of the national output; and productive services of economic resources exchanged in their respective markets at prices which constituted their incomes. In this system of an "autonomous, self-regulating" economy, the role of government was limited largely to the protection of property rights and persons, the administration of justice, the maintenance of internal order, and the security of a nation's frontiers.

Laissez faire was dominant in Britain during the nineteenth century. It became the prevailing view in the American economy where new land, an abundance of other natural resources, and the absence of Old World restraints resulted in a highly developed individualism. Elsewhere in western and central Europe, and in other parts of the world having close economic and political ties with the Western world, economic liberalism also made headway. Attending it came an era of great industrial expansion and rising standards of material well-being never before witnessed in history.

Postliberalism

Economic liberalism in its "pure" form had a short period of dominance during the nineteenth and early twentieth centuries. Although in the United States it remains dominant as a social philosophy and as a way of organizing and regulating the economic processes, its force has waned. Elsewhere its grip has weakened, or has been broken entirely, as national governments have increasingly established control over the economic processes. Liberalism in its origin had asserted the right of the individual against the outmoded traditions and governmental privileges and regulations thwarting his free choice and the pursuit of his self-interest. It protected him from the tyranny of government and the entrenched interests and "gave him his head." As time went on, however, the *laissez-faire* policy was criticized for failing to protect the individual sufficiently against the economic and social dislocations that at least in part resulted from the free play of self-interest in economic affairs. Individualism places a premium on initiative and ability in money-making pursuits. Since all people are not equally endowed

with this type of ability, or do not have equal opportunity to get ahead in the economic struggle, economic liberalism produced its own distinctions—based on money—between poverty, comfort, and great wealth. The economically weak fared badly in the competitive struggle, where each looked out for his own interest. This was particularly evident in the early phase of industrial capitalism during the nineteenth century. Wage earners worked long hours for low wages under oppressive working conditions; child labor was widespread. Although the "laboring poor" were undoubtedly better off materially than were those of some earlier periods, the growing humanitarianism of the nineteenth century recoiled at the human degeneration prevailing in the industrial slum areas. Positive interference by government was demanded by the reform groups in most industrial countries.

Among the early departures from *laissez faire*, therefore, were those laws designed to remedy these social ills. Child labor was curbed, working hours were shortened, prevention of industrial accidents was instituted. In the course of time legislation also extended widely to protect groups who, in the normal working of economic self-interest, are unable to attain standards deemed suitable to society. It should be emphasized that this legislation did not attack economic liberalism per se but was advanced as a means of attaining social welfare; governmental intervention was sought as a supplemental means of attaining this accepted end.

The policy of *laissez faire* has been weakened considerably by special groups within the economy who seek protection. In this country manufacturing groups and organized labor at an early date sought tariff laws to protect them from the competition of foreign goods. Small business has since the end of the nineteenth century sought government aid against big business and monopoly; and big business seeks special favors from government at all opportunities. For several generations farmers have demanded governmental aid on their behalf, such as a reform of the money system, the breakup of monopoly, the assurance of "fair" prices, subsidies, and cheap credit. Even the public in general may be considered, in the light of *laissez faire*, as a special group obtaining protection against the free operation of industry through such means as governmental control of public-utility rates and the imposition of standards on foods and drugs. Thus there has been a burgeoning of governmental intervention in behalf of farmers, business groups, and wage earners to protect their economic interests. The liberal faith in competition as the means of directing individual interest toward the public welfare has thus been discarded over a wide sphere of economic life. Government now regulates and controls an increasing measure of the economic processes.

The *laissez-faire* policy has been weakened in still another direction. A relatively free economic system is subject to periodic depressions which cause heavy losses in capital values, profits, wages, and in the volume of production. Here and in other countries governments are now wont to intervene to stabilize production, prices, and employment and to attempt to maintain "full employment." Under economic liberalism such efforts were regarded as unnecessary, for it was believed the normal operations of a competitive economy would bring about relatively full levels of employment and production.

In other ways the *laissez-faire* policy has been criticized and in part abandoned. Self-interest operating through private enterprise is expected to produce the things that the consuming public wants and is able to buy. Consumer purchases depend upon (1) individual incomes, (2) individual preferences or wants, and (3) the prices of goods and services. This, it has been said, is the way consumption is regulated and limited in a system of economic liberalism. Consumption on the part of people with low incomes is rigidly limited by their restricted capacity to spend. Accordingly, government has stepped in to provide many things, especially services, so that those who are unable to buy such services may have them free or at a price below that charged by private producers. Cases in point are public education in this country and elsewhere, the rendering of medical services to the poor, and governmental low-cost housing. Also, government indirectly affects consumption by aiding low-income groups through minimum-wage legislation, social security benefits, its encouragement of labor unions, and the payment of subsidies to farmers. These various policies alter income distribution and consumption to some extent, and therefore the composition of the national output.

In many ways the era of postliberalism is akin to mercantilism. That is, self-interest alone is not trusted to guide the processes of production, income distribution, and consumption. The harmony of individual and social interests is denied over a wide range of activities. The operation of the economy has become more and more a "responsibility" of government, under the "dextrous management of a skillful politician" and his economic "planning" experts. The resulting system is sometimes termed a *mixed economy* or *regulated capitalism.* It is mixed in the sense that private interest—the decisions of private enterprise and consumers—is modified, regulated, and supplemented by governmental decisions.

Governmental intervention has proceeded unevenly in the nations that had embraced the policies and practices of economic liberalism. In Europe, two world wars, revolutions, and widespread depression have caused the

restoration of a large share of authoritarian control over the economic processes. The British economy has become socialist. The economy of the North American continent (excepting Mexico) remains largely "free" in the liberal sense although not without fundamental changes in the relation of business to public authority which may possibly lead to great restrictions in the future.

THE CONDITIONS OF ECONOMIC PROGRESS

After the decline of feudalism and its static conception of social affairs there was ushered in a long period of economic expansion during which national outputs increased, first in western Europe and England, later and most spectacularly in America. An increase in output rests upon improvements in technology and upon the *accumulation of capital,* which is discussed in the remainder of this chapter as the basis of the economic progress achieved in modern times.

For present purposes the main types of capital are: (1) *capital facilities,* such as tools, equipment, machines, ships, docks, other transport facilities, and buildings and installations for production, and (2) *the materials that go into the making of finished goods.* The increase in capital from the feudal era to the present occurred first in the sphere of trade, particularly international trade, then in manufacturing and agriculture, and in more recent decades in the service industries. Of relatively small importance in feudal economy, capital has emerged to command the modern productive process. The conditions conducive to this growth may be treated briefly under two headings: the psychological and the objective conditions essential or desirable to the accumulation of capital.

The psychological condition refers merely to the outlook and spirit of a group, or a community, or a people. Willingness to work, save, and invest is essential to capital accumulation; a spirit of enterprise is necessary to spur on the productive activity of the economy. The feudal spirit was unsuited to the accumulation of capital for productive purposes; the medieval idea of "blessedness of poverty" gave neither the individual himself nor the community the impetus to economic expansion. The greater part of such capital building as occurred was devoted to the purposes of the Church and of defense. Materialism became a central concept in the modern age along with rationalism and science. Thus the psychological condition became such that capital accumulation was both a political goal and an object of individual striving, receiving the sanction of the prevailing social philosophy.

10453

This condition does not presuppose that all persons or even most persons in a community or nation must be interested in capital accumulation. Undoubtedly the bulk of the population in any age remains indifferent; in the past people's interest has largely tended to center on consumption and leisure rather than on capital accumulation. Indeed in the mercantilist economy both their leisure and consumption had to be curbed in order to expand production and assure the excess of production over consumption—capital accumulation. It was sufficient that a relatively small group had both the ability and the will to save and invest and thus bring about the accumulation of capital.

Given the favorable psychological condition, objective factors affect the form and the extent of capital accumulation. These include an adequate and industrious labor force, a talent for organization, suitable and easily accessible natural resources, an advancing technology, a favorable public policy, good markets for the sale of products, a distribution of income favorable to capital accumulation, and other factors of lesser importance. From the early nineteenth century these conditions existed in the United States.

Growth of the American Economy

When the United States embarked upon its national career the economy of the country was in an earlier stage of development than the economies of the powers across the Atlantic. In population it was less than four million, or about the size of Portugal or the Netherlands. Over 90 per cent of the population lived in villages or on farms; only 3.3 per cent lived in the six towns that had 8,000 or more population. Jefferson predicted it would take a thousand years for the wilderness as far west as the Mississippi to be thickly settled. Trade, however, was profitable and grew rapidly; exports were chiefly raw materials, and imports were almost wholly manufactured goods. Manufacturing had hardly begun by 1790; the value of manufactures at this time was estimated at 20 million dollars—which equals the annual output of only a single medium-sized industrial firm today. Seventeen branches of manufacturing were reported in existence, chief of which was shipbuilding. Paper, printing, leather goods, glassware, furniture, liquor, sugar refining, textiles, and iron making existed on a small scale. The paucity of working capital, difficulties in getting machinery, shortage of skilled workmen, a poor currency, transport difficulties, and the preponderance and attraction of agriculture—all limited the development of industry.

From this uncertain beginning, the economy of the United States grew to its present position as the leading industrial system of the world.

The existence of enormous quantities of readily available *natural resources* contributed greatly to growth in national output. Land and timber resources were plentiful, coal and iron ore were of high quality and suitably located. Petroleum deposits were rich, and water resources were adequate for both power and convenient transportation. Both climate and soil favored a varied and productive agriculture. In short, the physical setting was highly propitious; land, energy, metals, petroleum, food, and raw materials—the main ingredients of the productive process—were available at low cost. Their availability, along with the vast outpouring of human energy, go far toward explaining the more rapid advances in American production compared with other nations less liberally endowed by nature.

During the course of the nineteenth century *geographical expansion* westward brought more of these natural resources within the orbit of the economy. Beyond this factor geographical expansion is of itself stimulative, for it requires the building of transport facilities, of cities and towns; it creates opportunities for productive enterprise. The expansionist policies of the government, from the time of the Louisiana Purchase to the closing of the frontier in 1893, coupled with its liberal policy of disposing of land and other resources, were major stimulants to the economy.

Population growth was another factor contributing to American economic expansion. From 6 million in 1800 the population rose to 76 million in 1900—nearly a doubling every 25 years. A high birth rate and large immigration accounted for the rapid expansion during the last century. This growth was stimulative for several reasons. It provided the growing labor supply needed in the new country. Moreover, the additional population required housing, food, clothing, and innumerable other commodities and services; it created a *potential* demand for these things, and increased employment transformed potential into effective demand. In still another way stimulation resulted. A rapidly growing population is "young," i.e., a large proportion of the total population is composed of youths and young men and women. It is from young groups that initiative and enterprise flow freely; the tasks of pioneering in farming, industry, trade, and finance, for example, are more readily undertaken by the young than by those who have been slowed by age.

Production per capita or per worker since the early nineteenth century has increased steadily. This rapid rise cannot be accounted for by increases in resource use and population alone. *Technological innovations and improvements* constitute another important factor. The nineteenth century

brought forth a veritable technological revolution in both the United States and the industrial nations of western Europe. Techniques in practically all fields of manufacturing were transformed by steam power and later by electricity. New metallurgical processes provided the cheap metals required for the machinery, tools, and utensils. From the handicraft stage of production prevailing in 1800, this nation emerged by the end of the century as a leading industrial economy; today it is the world's greatest industrial power. Perhaps most vital to this change were the developments in the technology of transportation; the great areas of the west were linked economically to the industrial east by the rapid spread of railroads. Agricultural techniques were also greatly affected by the industrial development; the mower, reaper, thresher, and a long series of mechanical aids enabled vast new areas to be cultivated with but a small amount of labor. Furthermore, technology created new products and new industries; and many old industries were modified. One new industry after another appeared on the industrial scene, each adding its stimulative influence to the economy as a whole. Changing technology in all these phases has been a decisive, dynamic factor in the expansion of the American economy, encouraging the growth of capital and accelerating it by increasing the productivity of all factors of production.

American economic expansion owes much to the *policies of government.* In the first place, constitutional guarantees of personal and property rights firmly established the legal basis of economic individualism. Until recent decades these rights were interpreted in a sense which limited strictly the power of government to intervene in economic affairs. In other respects government directly encouraged and stimulated economic activity. At little or no cost government turned over most of its vast public domain—land, timber, water, and mineral resources—for private development and exploitation. The importance of this liberal policy cannot be exaggerated; government "spent" its natural resources freely, to the great benefit of industry and agriculture. By erecting tariff barriers it protected industry; through its monetary and banking policies it established in the course of time a stable nation-wide financial system. Because government was limited in scope, taxes on business and individuals were of little importance in the period of greatest expansion. In general, governmental policy and sentiment have traditionally given individualism wide scope to develop and expand the productive powers of the economy. With large amounts of cheap natural resources, a progressive technology, and the blessing of government, private enterprise has had in the United States its golden opportunity, literally and figuratively.

These favorable conditions were accompanied by a *distribution of income* conducive to saving and capital growth. The relationship of income distribution to economic expansion may be stated as follows: total income of the economy is equal to output.[8] If all income were spent for consumption purposes, the resources of the economy would be devoted entirely to the production of consumer goods,[9] and capital could not be accumulated. This is why saving is essential to capital growth. When part of total income is saved, some economic resources are freed from producing consumer goods and therefore may be devoted to the production of capital goods. Income inequality encourages savings. The wealthy normally save a large part of their high incomes, while the medium- and low-income groups usually spend all or almost all of their incomes for consumption purposes. From the evidence at hand, income inequality was pronounced in the period of rapid American economic growth. Relatively low incomes for the bulk of the population and substantial savings on the part of the wealthy kept consumption sufficiently below total output to enable a large accumulation of capital to take place. Had incomes been less unequally divided, had the low incomes been somewhat higher and the high incomes considerably lower, savings and therefore capital investment and economic expansion undoubtedly would have been less than in fact they were.[10] Although in its early stages industrial development is furthered by relatively low levels of consumption, conditions differ in a more fully developed industrial system; in the latter stage relative high levels of consumption are both possible and necessary for continued growth.

These factors have been described here: large quantities of natural resources, geographical expansion, growth of population, technological innovation and the appearance of new industries, governmental policies favorable to individual initiative, and the distribution of income. They are among the objective influences that were largely the source of American economic expansion throughout the nineteenth century. Conditions in recent decades have changed in important respects; these changes will be discussed in Chapter 7 as they relate to the prospects of future economic

[8] In Chap. 4 the relations between income and output will be explored in some detail. For present purposes to say that income equals output is a useful simplification.

[9] Except for the production of capital facilities to replace those that wear out.

[10] Capital accumulation as such does not necessarily require an unequal distribution of income. A socialist economy could hold all incomes relatively low in order to keep consumption well below output while the state "saves" and invests on a large scale. In Soviet Russia the incomes of the masses are kept very low thus allowing the state to save and invest in its industrialization program. However, income inequality is marked also in the Soviet Union. Both capitalism and socialism accumulate capital in the same way—by curtailing consumption below total output, but somewhat different techniques are employed to attain this purpose.

expansion. Yet it should be noted that the factors described here do not by themselves explain the phenomenon of growth. As has been stated, a spirit of enterprise, a dynamic initiative, is essential to far-reaching economic expansion, whether the economy is *laissez faire* or government-controlled. This is because investment in capital is necessarily risky, particularly when it takes the form of huge expensive capital facilities. Investment in capital on a large scale continues for many years, often several decades, into the future. A strong faith in the economy and a willingness to gamble personal gain characterize the making of such investment decisions. The dominant figures in past American economic history were imbued with such a spirit, and indeed often a ruthlessness, in the exploitation and development of economic opportunities.

Economic expansion in this country during the past century was a part of a general growth throughout the Western world. The elements of a new society that emerged from the ruins of the feudal age were given their most expansive development during the century preceding World War I. The period between world wars saw the establishment of authoritarian governments, a movement encouraged by depression and revolution; this movement has profoundly affected the levels of economic activity in the leading nations of the world, and its influence has supplanted in part the doctrines that shaped the growth of the major economic systems during the last century.

CONCLUSION

In this survey of the past five or six centuries two main issues have been taken up: (1) the organization of the principal economic processes and the significant changes that have occurred in their organization at various times, and (2) the conditions governing the level of economic activity as it has progressively increased in modern times. The first issue relates to the way in which a given complement of economic resources is allocated among the various uses for them, and the way in which the output produced by these resources is distributed, and finally, the regulation of the use or consumption of this output. These are the fundamental economic processes and have their origin in the scarcity of goods and resources in relation to demand. The second issue may be termed *economic progress*, i.e., the growth of national output and the conditions conducive thereto. These two issues are distinct but related. The material well-being of society is governed both by the organization of an economic system and by the level of employment of its resources.

PART II

NATIONAL INCOME AND THE EMPLOYMENT OF RESOURCES

CHAPTER 3

The Process of Production

The position and prestige of a nation in world affairs and the material well-being of its people are in large measure determined by its productive capacity and performance. The people of the United States have generally lived on a higher material plane than the peoples of Europe; western European standards have been higher than eastern; and Europe as a whole has exceeded the Orient in economic welfare. These variations in planes of living are largely caused by differences in productive systems and in per capita production. Material well-being in any country can improve only if production is increased, thus providing both the commodities and services people want and the money income with which to buy them.

For individuals, the term *standard of living* is often used to designate how much money income—in wages or salaries or dividends—they receive, and what they can buy with these incomes. People are paid because they produce; with their incomes they buy the things that they and other producers are turning out in the factories, farms, stores, and other productive establishments of the country. The amounts of income individuals receive and the amounts of things they buy with their incomes determine their standards of living. But on the whole what they receive as earners and buy as consumers depend upon the level of productive activity in the nation.

Production is a central interest in economics. The position of the nation in international affairs—both in war and between wars—is directly depend-

ent upon its productive capacity. During war, productive capacity enables a nation to engage in military operations while sustaining its civilian population. Modern war is a war of productive systems; the decision is won by that nation or group of nations able to produce the largest amount and most effective kinds of military supplies and equipment. Events of recent years leave little doubt on this score. Between wars the economic position of a nation exerts great influence on its relations with other parts of the world. A large industrial nation may produce commodities and perform services vital to the well-being, and perhaps the existence, of other countries. Its imports of raw materials and other commodities may help sustain the economic systems of countries in economically less dominant positions. An obvious example is Great Britain whose power and prestige during the nineteenth century were based mainly on its place as the foremost industrial nation of the world, as the "workshop of the world." The industrial growth of Germany late in the century threatened both the economic and political position of Great Britain. Similarly, the growth of Japan's power during the twentieth century—and its bid for dominance of the Far East—were made possible only by the expansion of its productive capacity.

No modern nation has attained pre-eminence as a political power without a solid base in productive capacity. Changes in the relative industrial positions of the nations of the world in the future will inevitably lead to the decline or rise of certain nations. If a nation, therefore, is to maintain or improve its position in world affairs, its productive capacity must keep pace with, or surpass, that of the leading contenders in the ceaseless international contest for position. Production alone, however, does not assure a place in international councils; it provides the material support of a nation's candidacy. It falls to diplomacy to use these means to best advantage.

Interest in maximizing a nation's production has its beginnings as far back as the seventeenth century, during the mercantilist period when economics first assumed its modern form and outlook. Its earliest expression was largely nationalist—the desire to increase the wealth and power of a nation at the expense of other nations. Adam Smith's monumental work, *An Enquiry into the Nature and Causes of the Wealth of Nations,* published in 1776 is an investigation of the conditions that increase productive capacity and a declaration that economic liberalism is the best means for attaining maximum output. Smith pleaded his policy to enhance the power and glory of Great Britain and, concomitantly, to improve the lot of its citizens. Economic liberalism established itself as an almost irrefutable doctrine when the prodigious expansion of production during the nineteenth

century, first in England, later in the United States, gave support to its thesis.

More recently economists have revived their interest in production—or in national income as the monetary value of productive output. This preoccupation was aroused by the prolonged depression of the 1930's, when a considerable part of the productive capacity here and in other countries was not used. The enormous increase in production (or national income) during the war and the problem of maintaining production and employment in the postwar years further stimulated general interest in the causes and conditions underlying the level of production. The attention of others in production extends to farther horizons. Will the expansion of American economic activity—and the strength and prestige that go with it—keep pace with that of Russia, of Germany, and of other industrial powers that may emerge? Are the conditions that made the United States industrially strong in the past capable of bringing about continued expansion? Are there conditions within this country that might impede its future growth? And, if so, what might be done to assure continued economic expansion?

ELEMENTS OF THE PRODUCTION PROBLEM

Before going into the discussion of the conditions governing the volume of production, the separate, but related, elements of the production problem should be distinguished. The economic resources available in a nation—such as human efforts and talent, capital, and natural resources—form the basis of its productive capacity. The size and quality of its labor force, its plant, machinery, and equipment, its land, minerals, energy, and other productive resources, establish the upper limit to a nation's productive capacity. A nation can increase its production only to the extent that its resources will permit. The volume of production, or national income, is governed by two major decisions in the use of resources:

1. *The extent to which resources are used: levels.* Production will be at a high level when resources are fully used, as during war; production will be at a low level when much of the nation's resources are permitted to remain idle.
2. *The way in which resources are used: allocations.* When labor and capital are allocated in the most economic manner, production will be high; when labor and capital are allocated between efficient and inefficient fields of activity, production will be relatively lower.[1]

[1] Considerable attention will be given in Chaps. 12-21 to the ways in which labor, capital, and other economic resources are allocated.

Full use of economic resources, therefore, does not in itself assure maximum production. *Maximum output is attained by the best allocation of resources when coupled with high levels of employment.*

THE PROCESS OF PRODUCTION

Economic life is an organization of producers to satisfy the wants of consumers. While this statement is generally correct, it should not be assumed that the conscious purpose behind producers' activities is necessarily the satisfaction of consumers' wants. *The immediate purpose of most producers in a capitalist economy is to obtain an income—wages and profits—for themselves.* To accomplish this purpose most of them produce things that result in the satisfaction of human wants. If wages, and particularly profits, are not likely to be earned, production will cease or be curtailed at the expense of satisfying consumer wants. Although it might be said that the ultimate social purpose of productive activity is the satisfaction of human wants, this social purpose forms no part of the immediate calculation of a producer who makes a decision to invest capital, hire workers, and produce goods. In other words, the existence of human wants as such does not assure the production of commodities and services. Most human wants in this country today are satisfied by commodities and services produced by business guided by profit incentives. Some wants, however, are satisfied outside the sphere of business activity. Governments render services without direct charge, such as public education, public health services, police and fire protection, etc. And many want-satisfying services are performed within the family; e.g., the preparation of meals, housework, and gardening.

The dependence of production on monetary incentives is one of the distinguishing features of this economic system. Both in Europe and the United States people have at one time or another lived largely by their own efforts, producing most of the things needed for subsistence. Earning money was subordinate to the main task of producing food and clothing for family needs. The surplus over and above family needs may have been sold for money, but more often it was traded directly for other things produced by neighbors. A production system may, like that of Soviet Russia, subordinate the making of money income by individual producers to the production of goods in accordance with state purposes. These purposes obviously include the satisfaction of consumer wants, as well as state programs, such as military preparation, industrial diversification and expansion, and perhaps the maintenance of position and perquisites of the ruling bureaucracy.

The Complexity of Production

The fact that the productive processes in the United States are highly specialized and are organized chiefly on the basis of making money for the producers accounts for much of the complexity of the system. Human wants act as a general guide to production; they are manifested by the ability and willingness of people to buy things. Producers must calculate this willingness in advance of production, and to make money they must produce those things people want badly enough to pay more for than they would for other things that could be produced. This is an enormously complex process, for there are often many producers catering to the same wants, and many wants are unstable. Since the satisfaction of wants depends largely on purchasing power or money income, the producer's success in making money also depends on his prospective buyers having an income and turning over part of it to him for the things he produces. *Thus, wants on the one hand and producers' decisions to turn out goods on the other are not directly related; they are, rather, indirectly related by prices, costs, money income, and markets.*

Production is a process that involves a continuous *input* of resources—capital, raw materials, labor, and managerial effort—and an *output* of commodities and services in a form useful to the final buyer, or useful to other producers in further stages of production directed toward the same end. Almost any commodity in common use is the product of a long series of diverse and complicated productive stages.

For example, consider the production of a can of string beans. In their raw form the string beans were grown in New York state, where farm labor was hired to plant and harvest the crop. The beans were shipped by truck to a near-by cannery, where other labor and a fairly elaborate set of machinery processed the beans and canned them. From there a railroad carried the can to a wholesaler's warehouse, where it was held for several months, before being shipped by truck to a grocery store, to be sold by the grocer and delivered by a grocery boy to the final consumer. The last act in this long process is the preparation of the string beans, together with other foods with perhaps an equally complicated background, for table consumption. These stages from the farm to the dinner table required the services of farm labor, of truckers and railroaders, the canner, the wholesaler, and, finally, the retailer. For cultivating and harvesting the farmer used machinery made in Illinois; the steel in these machines was milled in Pennsylvania; the coal to make the steel was mined in West Virginia, the iron

ore in Minnesota. Ships and railroads brought these materials from their sources to the fabricating centers. The cannery was built from lumber cut in Oregon and shipped by vessel to New York; the cannery machinery was made in Pennsylvania, the cans in Ohio, of tin from Bolivia smelted in Great Britain. The trucks, locomotives, and cars, the warehouse and retail store were produced by labor and materials from all parts of the country. Common labor, skilled labor, bankers, insurance men, salesmen—all had some part directly and indirectly in bringing the string beans finally to the dinner table. Mines and factories in all parts of the country and in foreign countries contributed indirectly to the satisfaction of eating the string beans.

All these activities are *productive* whether they produce directly the goods consumed, the string beans in this example, or the goods essential to their eventual consumption: steel, tin, coal, lumber. The services performed by the railroaders, the truckers, the wholesaler, the retail clerk, the delivery boy, the banker who financed the canner and wholesaler, the insurance man who insured the string beans in their journey from the farm to the retailer—all are productive. The services of the post office or the telegraph company in transmitting the orders for the materials involved, the services of the schools that trained the workers, the services of the police and fire departments who stood ready to protect the string beans during all stages—these are also productive.

In these operations, economic resources are put into use, and from this use the thing in production changes its form (from raw to processed string beans), or its location (from the farm to consumer's home), or it is held (by wholesaler and retailer) until needed and finally sold to the consumer. As noted in an earlier chapter this is the creation of *utility*. The input of resources requires an expenditure of money to hire, rent, or buy the various resources used: these are the money costs of producing the final product, which may be incurred long in advance of final production of the commodity. The price of the final product is expected to cover the costs incurred in the process. These costs in turn are the wages, salaries, and other incomes which make up the purchasing power with which the product may be bought.

The many stages traced here give a typical pattern for the major part of the production process. It shows the vast specialization in production today. Specialization means that producers in general work for somebody else. Wage and salary employees work for an employer, but the employer works for the buyers of his finished product, much as a dentist works for his patients. The only producer who "works for himself" is the mountaineer who scratches his hillside to grow his food, weaves cloth from the wool

of his sheep, makes his own garments, and otherwise provides for no needs but those of his family. This man is independent, but quite likely very poor. Specialized production is more efficient.

Specialization, it is obvious, requires an exchange of goods. Exceptions to this principle are those goods and services produced in the household for household use, and many of the services of government (such as police and fire protection) which, although part of the specialized production process, are not sold in the usual manner. But the bulk of the nation's productive activities are conducted for the purpose of exchange, or sale.

This point is of great importance since the level of output and the types of goods produced are governed by the prospects of advantageous exchange. Prospects of sale exert a selective influence on production, encouraging those types of goods and services which promise a good financial return and discouraging those that do not. Thus, whether production in general is running close to the capacity of the nation's resources, or at some point below this level, depends largely upon the prospective sales and financial returns to the producers. Sales depend on the demand of consumers for goods and services and the demands of business and government for materials, equipment, and machinery necessary to maintain current production and to expand productive capacity. The demand of consumers is largely determined by the income they receive in the form of wages, salaries, or other payments. The demand of business for materials, equipment, machinery is governed by present and prospective profits, which are ultimately determined chiefly by the demand of the consuming public. The cycle of production and exchange, then, is an outflow of goods and services, an outflow of incomes as the cost of these goods and services, and the return of these incomes to buy the goods and services produced for sale (exchange).

ECONOMIC RESOURCES

Meaning of the Term

What is produced and how much can be produced are matters that are partly determined by the availability of resources. Obviously, metals cannot be produced when ore is nonexistent, inaccessible, or too poor in quality to warrant mining operations; crops cannot be produced from soil and climate entirely unsuited to their growth. Nor can any amount of commodities be produced at will—production is kept within the limits of existing resources.

Broadly defined, *economic resources are those elements in the environment that are useful to man, that have a capacity to satisfy his wants directly or indirectly.* These elements are resources only when they serve man to satisfy his wants. Coal in the Pennsylvania hills was not a resource for the Indians who originally inhabited the region, but it became so for the settlers who had the knowledge and means of using it. If coal were completely displaced by other sources of energy, such as atomic energy, it would lose its resource qualities. The variety of resources at hand is partly determined by differences in natural environment and partly by differences in the character of society. The physical environment of the area of the United States was about the same four hundred years ago as it is now; yet the resources of the 1540's were substantially different from those of the 1940's. A hunting and pastoral society prized certain elements which in an industrial society might be valued chiefly for recreational purposes, while coal, oil, and metal ores cherished today were either hidden from or ignored by inhabitants four centuries ago. The waterfall that required an earlier man to expend energy in portage is a source of energy for modern man.

Economic Resources as Factors of Production

The elements in the natural and social environment that function in the service of man for the satisfaction of wants are termed the *factors of production.* These are customarily classed as land, capital, labor, and enterprise (management). By "land," economists mean natural resources or natural agents. Land was classed in the seventeenth century together with labor as the base of production. "Labor is the Father and active principle of Wealth, as Lands are the Mother," according to Sir William Petty. Land was formalized by the early nineteenth-century economists as one of the four factors of production, and was considered the most important natural resource, indeed, the greatest form of wealth. Land, however, no longer holds such an exalted position; ore deposits and energy sources have become under industrialism a contender for first place among the natural resources. It is therefore inappropriate at this point to use the word "land" in its traditionally inclusive sense.

Economic resources then may be classified as follows:

A. *Natural resources* including:
   1. *Land* or the climate-soil-topography combination
   2. *Forest resources*
   3. *Minerals,* or the fuel-metal combinations

4. *Water resources,* including water as a source of power, as an avenue of transport, and as a source of food.

B. *Capital resources,* or the produced goods used in the process of production. The tools, machines, equipment, buildings, improvements, and installations used in the growing, extraction, fabrication, transportation, and final sale of goods are the important forms of capital. Often, capital is virtually merged with the assets of nature itself, as when hydroelectric, irrigation, and reclamation projects are built and become part of the natural environment. Materials in process of production and finished goods found in manufacturers', wholesalers', and retailers' inventories are also part of capital. In this book, the discussion of capital will relate largely to *capital facilities,* i.e., tools, machines, equipment, buildings, and installations.

C. *Human resources,* or the efforts and talents of man essential to carrying out productive functions. The functions performed by man in production are:

1. *Enterprise.* This is the task of initiating projects and new productive developments; it involves the imagination, energy, and talent "to get things going" and to formulate policies and the principal decisions for the continuation and direction of productive activity.
2. *Organization.* This is the function of bringing economic resources together in an efficient combination and of establishing procedures and organizational forms to carry out projects conceived by enterprise.
3. *Management.* This is the work of executing policies and directing resources brought together in a productive unit.
4. *Risk-bearing.* A productive enterprise usually requires the venturing of time, effort, and money on the part of some without assurance of future payment. Uncertainty of reward is characteristic of productive activities catering to future wants.
5. *Labor.* This includes the great variety of tasks both manual and mental carried out under the direction of managers.

These various functions may be carried out by one person. For example, the individual entrepreneur may initiate a productive undertaking, organize and manage it, risk his own funds, and do some or all of the work required. In the corporate form of business organization these functions tend to be separated in various ways. The term *top management* is often applied to

corporate officials who initiate work, formulate policies, and make the principal decisions; below this level is the management group that carries out policy and directs the work of wage and salary employees. Financial risks are borne by officials and employees to the extent that they own stock in the corporation; in the large corporation, however, much of the financial risk is borne by stockholders who usually perform no other function than bearing some risk.

These resources vary greatly in quantity, quality, and types; their world distribution is uneven. The most advantageous climate-soil-topography combination for agriculture and animal husbandry, coupled with a relatively advanced use of capital in the production of food, is found in the United States, Canada, Great Britain, and northern Europe. Similarly the major combinations of fuel-power-metal are concentrated in the North Atlantic region. In these areas too are found the human resources which because of ethnic characteristics, climate, and cultural conditioning are most adaptable to the techniques of modern industry. This North Atlantic region is the seat of the dynasty of coal and iron, mechanical energy and metals, where the bulk of world industrial production and much of its food production are located. Here is the great surplus area, producing far more than necessary to sustain its populations, while Asia, with two thirds of the world's population, gets along with little more than enough to feed its people.

ECONOMIC RESOURCES OF THE UNITED STATES

The United States, with its varied natural environment, combines and co-ordinates agriculture, manufacturing, mining, and the service trades to a degree not found elsewhere.[2] Other countries less liberally endowed with land, forests, and minerals have developed in significantly different ways. In Germany the scarcity of resources impels a more careful and intensive use of them than is customary in this country. Labor, however, has always been more scarce here than in most other parts of the world; consequently the United States leads the world in the conservation of labor resources. It will be seen in the discussion following that the way in which economic resources are combined and used is governed in large measure by their relative scarcities.

[2] The service trades include transportation, communications, entertainment, hotel, and innumerable others that render services to consumers and industry. The service industries may be contrasted with "commodity-producing" industries.

Natural Resources—Land

Land resources—a combination of climate-soil-topography—are the bases of agriculture and animal husbandry. Of the 1,903,000,000 acres in this country, one sixth (350,000,000 to 380,000,000 acres) is normally in crops. Two thirds of this total is devoted to production of feed for livestock, one fifth for food production, and the balance for cotton, tobacco, and other inedible raw materials. Some 800,000,000 acres are used for pasture. Thus, most of the nearly 1,200,000,000 acres in agricultural and livestock use are devoted to feed-growing and pasturage; only a small part is in food-crop production. These figures indicate a large per capita supply of land, a condition which must exist if a nation is to devote so much of its crop land to livestock feed and so much of its total acreage to pasture. In densely populated areas most of the crop acreage is in foodstuffs, pasturage being only a small percentage of the total acreage. India and Japan devote almost all of their crop acreage to foodstuffs; consequently animal products are relatively unimportant in the national diet (in contrast to 40 per cent in the American diet).

The large supply of land in this country has since the beginning of settlement been accompanied by a shortage of labor. Consequently labor has been used sparingly as a general rule; the typical labor-land combination has been much land and little labor. One result of this shortage has been a tendency to "mine" the land, to exhaust the fertility of a plot and abandon it for new ground. This practice was actually the most economical under past circumstances, because a careful use of the land would have entailed the employment of more labor per acre, thus raising the cost of producing foodstuffs and raw materials. Land would have been "conserved," but labor would have been "wasted," i.e., used in a less productive employment. Labor, being more scarce, was relatively more important. Undoubtedly this practice contributed to the rapid utilization of land areas in the United States, but it also caused the ruin of a part of its natural resources.

The plenitude of land and the scarcity of labor led early to an emphasis on the use of capital in American agriculture. Mechanized agriculture is best adapted to extensive acreages. The degree of agricultural mechanization varies considerably in different parts of the United States, depending as it does on topographical, crop, and labor-supply factors. The plains area is most suitable geographically for mechanization, and high labor costs (scarcity of labor) encourage the adoption of machine methods in this region. On the other hand, the southern states having a less scarce and

therefore cheaper labor supply, have made more economical use of resources by employing labor rather than machines on a large scale. However, the recent growth of manufacturing in the South with the consequent shift of population to cities makes farm labor scarcer and therefore more costly than in the past.

Other Natural Resources

Forest resources in the United States are extensive and generally accessible for exploitation. *Energy* resources, such as coal, oil, and rivers for the generation of electric power, have provided the basic energy for industrial production. Their proximity to raw materials, such as *metal* resources, and to transportation, such as *water* transit, encouraged an extensive and rapid development of these energy resources. Navigable rivers and long coast lines offer cheap transport facilities; and the topography of the country placed no major obstacles to the extension of rail transport linking up the raw material producing areas, the manufacturing centers, and the markets for finished commodities.

The past use of many of these natural resources has been similar to that of land. In many areas forests have been cut down at a rapid rate, but relatively little attempt has been made at reforestation, or careful selection of trees for felling. The scarcity of labor made it immediately uneconomical (unprofitable) for the producers to use labor to maintain forest resources when more could be found on the next ridge of mountains. Mineral resources have been exploited in the same general way, because this method of using plentiful and cheap resources was more economical in the short run than any other. In contrast, the mineral and forest resources of European countries are very scarce, while labor is relatively cheap, so that labor is used to a fuller extent for conservation.

Use and Conservation of Natural Resources

In one sense the rapid exploitation and lack of conservation of natural resources in the United States is wasteful: materials are lost. But a different use of economic resources probably would have meant less production in the past. There is, however, another aspect of this question. Some natural resources are replaceable, others are irreplaceable. Forest resources are replaceable over a considerable period of time; land resources are replaceable, i.e., fertility of the soil can be maintained, if proper farming methods

are used. But minerals generally are irreplaceable; deposits are eventually exhausted. The rapid drain on petroleum reserves, particularly during the war, has led some geologists to predict a serious shortage in two or three decades. What will a future generation use as a source of energy when oil and perhaps coal are exhausted? Some advocates of conservation propose a curtailment in the use of oil and other exhaustible resources; others propose a more careful use of these resources. Either course would defer some present uses of these resources to future uses; immediate profit of the individual producer would yield to the material benefits of future generations; present levels of production and national income would tend to be restrained in the interest of future levels of production and income.

A nation has a vital interest in increasing its future productive capacity, which under industrialism is based chiefly on energy and metal resources. The important sources of energy in this country—coal and oil—are exhaustible. Metal deposits, too, are exhaustible. The continued depletion of these resources will eventually destroy the production system unless other energy sources, such as water power for the generation of electricity or atomic energy, and other metals, such as aluminum and magnesium, are adapted to production and utilized on a large scale. New resources and new technology might therefore avert the decline of productive capacity in the modern industrial empires. Should this eventuate, the exhaustion of many presently vital natural resources might be offset by an effective use of these resources. But no one can be certain that new resources will be found to replace those now being depleted. The problem of conservation is vividly posed by modern warfare. Prodigious amounts of energy and metal resources are consumed, so that the nation which is most successful is also most extravagant in using resources; future productive capacity is, in this sense, sacrificed for immediate gain.

Capital

The term *capital* has a number of meanings. Professor Hicks, the noted English economist, defines as capital "all those goods, existing at a particular time, which can be used in any way so as to satisfy wants during the subsequent period." [3] This definition includes natural resources as capital, for these resources can be used in future times to satisfy wants. Similarly, consumer goods held in stock or inventories must be included. A more restricted and for present purposes more satisfactory definition limits capital

[3] J. R. Hicks, *The Social Framework*, Oxford, 1942.

to *produced goods used in further production.*[4] The terms *producer goods* or *investment goods,* discussed earlier in the book, cover this category. The principal interest here is in one aspect of investment goods about which there is no disagreement in definition, namely, *capital facilities.* Capital facilities are largely machinery, equipment, buildings, and installations made by man and used in the productive process. This term is more limited than investment or producer goods, which are both the materials (finished, semifinished, and raw) used in production and capital facilities. These terms shall be used frequently, although the emphasis will be placed on capital facilities as one of the elements of a nation's productive capacity. Some of the main forms of capital facilities may be listed to indicate the scope of the term:

1. Building and installations (plant)
   - Factories
   - Mills
   - Foundries
   - Warehouses
   - Barns and sheds
   - Fences, platforms, tanks
   - Retail stores
   - Garages
   - Grain elevators
   - Commercial buildings
   - Hotels
   - Apartment houses
   - Hospitals
   - Post offices, schools, and other public buildings
2. Industrial equipment
   - Engines
   - Motors
   - Derricks
   - Furnaces
   - Boilers
   - Presses
   - Die-making equipment
   - Drills
   - Cutting tools and machines
   - Lathes
   - Jigs
   - Refineries
   - Mining equipment
   - Cranes
   - Condensers
   - Grinders
   - Turbines
   - Welding equipment
   - Distilling machinery
   - Distributors
   - Driers
   - Evaporators
   - Feeders
   - Finishing machinery
   - Forging equipment
   - Hydraulic equipment
   - Loaders
   - Machine tools
   - Office equipment
   - Pumps

[4] When a businessman says that he is "unable to get capital" he usually refers to his inability to borrow money from his bank or from another lender. The businessman wants money, but he usually wants it to construct a building or buy materials or machines—these items are capital. Thus money is generally regarded as a mere *claim* on capital and other goods, not capital itself.

Chemical equipment
Electrical equipment
Coolers
Cookers
Crushers
Rollers
Smelters
Tanks
Testing machines
Generators

3. Transport equipment and facilities
Locomotives
Railroad cars
Railroad tracks
Subways
Streetcars
Trucks and trailers
Ships
Paving equipment
Road-making equipment
Bridges
Tunnels
Buses
Terminals
Airplanes
Airports
Hangars
Highways and roads
Docks
Pipelines
Canals
Dredges
Barges

4. Communications and power
Radio stations
Power lines
Telegraph and telephone lines
Hydroelectric projects

5. Farm machinery and equipment
Tractors
Plows
Harvesters
Combines
Disks
Cultivators
Hay balers
Windmills
Corn pickers
Potato diggers
Poultry equipment
Milking machines
Cream separators
Spraying equipment
Irrigation machinery

This list is by no means exhaustive. There are thousands of items in the capital-facilities category.

The American economy, which is noted for its reliance on capital facilities, owes its indirect, complicated production process and great productive capacity to the large amount and variety of capital facilities. The National Resources Committee [5] estimated the value of nonagricultural plant and equipment in 1935 at 105 billion dollars, agricultural plant and equipment at 9 billion dollars, governmental plant and equipment 19 billion dollars, a total of 133 billion dollars in capital facilities. Estimates of this sort are necessarily approximate, but they show the magnitude of capital facilities in this country. The value of these facilities, considerably increased

[5] National Resources Committee, *Structure of the American Economy*, Pt. I, United States Government Printing Office, 1939.

since 1935, is about equal to the value of all goods and services produced by the economic system in a prosperous year.

Capital Facilities and Investment—Net and Gross

The production or building of capital facilities is known as "investment." Sometimes the producer makes these facilities for his own use, but more often he makes them for sale to other producers: both processes are investment. This use of the term investment—the making of capital facilities—is more limited than the customary usage. When a person buys a bond or a corporation stock he usually calls this an investment. However, in the sense used in this book and found widely in economic literature, *investment is that part of total production composed of capital facilities and other investment goods.* As stated in an earlier chapter, investment involves saving, or waiting, because the production of facilities is a diversion of resources from the immediate production of consumer goods.

With the passage of time and with use capital facilities wear out or *depreciate.* To maintain the stock of capital facilities a certain amount of investment is necessary annually. This is *replacement* of capital. If investment exceeds the amount necessary for replacement, the nation's capital facilities are expanded; this addition is *net* investment. Total or *gross* investment during a year is the investment needed to replace worn-out or used-up capital facilities and other capital goods plus net investment.[6]

Investment and National Output

Gross investment in capital facilities is part of total production during, for instance, a year. The remainder of total production is chiefly for consumption purposes, or for the immediate satisfaction of wants.[7] Thus total production or *gross national output* equals gross investment plus the commodities and services for consumption purposes.[8] *Net national output* equals net investment plus the commodities and services produced for consumption purposes. These terms will be discussed more fully in the following chapter.

[6] Although emphasis here is placed on capital facilities as a component of annual investment, it should be noted that changes in inventories or stocks of goods constitute additions to or deductions from current investment. Residential construction is often counted as investment. International trade balances, too, enter into a nation's annual investment; this aspect of investment is treated in Chaps. 34, 35, and 36.

[7] Part of national output takes the form of war goods.

[8] Plus war goods production.

It will be noticed that investment in capital facilities and other forms of capital represents an excess of production over consumption. A country without such an excess cannot add to its capital or maintain its capital intact. India and China produce little more than is needed for basic consumption by their populations and are consequently unable to increase capital facilities to any substantial extent. It requires the great bulk of the working population of China and India to produce foodstuffs which, indeed, constitute a poor fare. The surplus of production over consumption in the United States may be ascribed chiefly to the large supply of natural resources, especially of energy resources, such as coal, petroleum, and water power. This excess, as capital, has made possible constantly increased productive activity and, therefore, an ever-increasing surplus over consumption. At the same time the progress of technology increased the productivity of the additions to capital facilities. As a consequence the productive capacity of the American economy is capable of providing high standards of consumption and high levels of investment, when both are compared on a per capita basis with other nations. Further increases in this productive capacity, then, require the production of surpluses, in the sense used here, sufficient (1) to make up the depreciation of existing capital facilities, and (2) to allow for some net investment as well.

Capital Facilities and Natural Resources

Capital facilities and natural resources comprise the physical basis of a nation's productive capacity. In some respects capital facilities and natural resources are similar, in other respects wholly dissimilar. As stated earlier, some natural resources are replaceable (soil fertility), others are replaceable only over a long period of time (forest resources), and others are irreplaceable (mineral deposits). Capital is a replaceable resource. Although it wears out or depreciates, replacement is possible if there is an excess of production over consumption for this purpose. Capital is similar to land, in this respect, although for most practical purposes land is relatively fixed in quantity while capital can be increased by net investment. Capital, then, is an expansible resource. It is therefore dissimilar to mineral resources. In fact, as capital facilities expand, mineral resources are depleted because industrial capital draws on metal for its machines (as well as for goods made by them), and on coal and oil for power. During the last two decades the growth and use of capital facilities have probably drained more irreplaceable natural resources from the earth than were used during all previous history.

Capital Facilities and the Scarcity of Labor

The prolific natural environment of this country made the growth of capital possible; the shortage of labor made it necessary and economically worth while. American economic history has been characterized by a growth of laborsaving machinery and equipment. The large amounts of land and the scarcity of labor necessitated the use of capital equipment, i.e., farm equipment and machinery. In densely populated countries the existence of cheap labor makes it unprofitable to use much capital. The average production per acre is probably higher in Denmark than in the United States, but the average production per farm worker is lower. The use of capital to offset scarce labor here and the use of labor to offset scarce capital there, may fairly characterize the difference. Industrial developments in the United States indicate that mechanization has constantly tended to be laborsaving, because capital is cheaper than labor for a great number of productive operations. This trend is cumulative in a sense. Laborsaving capital increases the productive output per worker. This increase is used partly for more consumption per worker (or per person) and partly for more surplus (saving) over consumption. From this surplus more capital facilities emerge and the expansion cycle can be repeated. The older industrial countries of Europe, being poorer than the United States in natural resources, have a smaller basis for capital growth, i.e., a smaller surplus over consumption needs; but their lesser scarcity of labor makes the need for capital relatively less urgent. India is a sharper contrast, for there the great amount of labor obstructs, on economic grounds, a substantial growth of capital. Why should a costly harvesting machine be used in farming when labor is plentiful at low wages, or why should a steamship company install expensive cranes on the docks when Indian stevedores unload ships at lower cost? Capital accumulation actually occurs in such countries, but cheap labor keeps it at a slow rate. In the United States, on the other hand, capital is relatively cheap, labor relatively costly.

Thus, the way economic resources are used in the productive process is largely governed by the relative scarcities of these resources and by the extent to which one can be substituted for another.

Human Resources or Labor Resources

Without the application of human effort natural resources would be of little service in satisfying the wants of man; without human effort capital

facilities would not exist. Effort makes useful the endowments of nature: it creates capital from them. The efficiency of human effort depends upon the quality and quantity of labor itself, and upon the quality and quantity of other economic resources. Concerning the quality of labor three factors have been suggested as of vital importance—ethnic characteristics and inheritance, health and energy, training and social environment.[9] The ability to learn and to carry on sustained and often monotonous effort are characteristics of great importance for the modern productive process; and these characteristics appear most pronounced among people of north and west European stock, including those of the British Isles. Climatic conditions play an important part in the health and energy of a people. The variable climate of the temperate zones is favorable to physical and mental energy and high standards of health. No major industrial area is found outside the temperate zones, partly because a greater standard of physical efficiency is found in regions of variable climate. The efficiency of human effort is governed, too, by training and social environment. The discipline essential for sustained effort is acquired partly, of course, by the necessity of earning a living, but also by the customs and habits in a community or nation. The environment of an industrial community is generally not conducive to sustained leisure; so strong are its environmental influences that work is preferred by most inhabitants to the boredom and social opprobrium attending idleness. By necessity and tradition they "practice the hard efficiency that this age asks for money."

Size and Distribution of the Labor Supply

The number and the relative quality of productive workers are basic elements in the productive capacity of a country. The Census reports decennially that so many people are "gainful workers"; in 1940 in the United States this was 53 million or 40 per cent of the total population; in 1947 the number had increased to over 60 million.[10] But in addition to gainful workers there are millions of others—especially housewives and children—who perform productive services, such as cooking, laundry, housework, gardening, etc. These must be counted as labor resources, but neither their

[9] E. Huntington and F. E. Williams, *Business Geography*, Wiley, 1922.

[10] The estimate for any particular year of necessity obscures the changes occurring from month to month in labor supply. For example, in January, 1946, the Census Bureau estimate is 59.5 million; for July, 1946, the figure is 62.8 million; for December, 60.3 million. Over 3 million persons—mostly students and housewives—enter the labor force for temporary summer work and withdraw during the fall and winter.

number nor the value of their productive effort can be measured accurately.

Accordingly attention here is confined to the gainful workers. These are (1) people who work for a money income (wages, salaries, fees, profits) and (2) those who are able and willing to work but are unemployed. The first category includes those who work for wages and salaries as well as those "independent enterprisers" who own and manage their own businesses. Table 1 shows the gainful workers of 1940 classified by types.

Table 1 **GAINFUL WORKERS, 1940**

Types of workers	*Men*	*Women*	*Total*
Proprietors and managers, nonfarm	3,500,000	300,000	3,800,000
Proprietors, farm	6,000,000	100,000	6,100,000
Professional	1,600,000	1,500,000	3,100,000
Skilled workers	6,900,000	100,000	7,000,000
Clerical workers	6,000,000	3,000,000	9,000,000
Semiskilled workers	6,200,000	2,800,000	9,000,000
Unskilled workers	11,500,000	3,500,000	15,000,000
Total	**41,700,000**	**11,300,000**	**53,000,000**

Approximately 20 per cent of the labor supply in 1940 was engaged in agriculture; 17 per cent in manufacturing; 10 per cent in construction; 20 per cent in trade and service; 18 per cent in transportation, mining, professional work, public service, and miscellaneous fields; and 15 per cent was unemployed. In 1946 less than 5 per cent of the 60 million labor supply was unemployed.

The distribution of labor resources is determined directly by (1) the varied demands of the economy for the things produced by labor and (2) the amounts and types of other resources. The first determinant applies to specialization of production and therefore labor skills; the second to the quantity and quality of other resources. For example, in the United States some 12 million people are able to produce virtually all the food and other agricultural products needed. If land resources were less plentiful, relatively more labor, perhaps twice as much, would be required to produce this agricultural output, thus reducing the amount of labor available for other fields of productive activity.

The distribution of labor resources is also influenced by custom and tradition, educational training, and individual abilities. Custom and tradition may keep some people in coal mining, for example, although in other fields

their efforts might be more productive. The needs of the economy for coal could conceivably be satisfied with fewer coal miners, but the remaining coal miners might be reluctant to change jobs, or may be unfitted for new tasks because of a lack of training or natural abilities required for new jobs. Educational opportunity and the ability to learn are important especially in defining the skilled groups and limiting shifts between types of specialized work. Unskilled workers, for instance, can scarcely become doctors or lawyers if they lack the natural abilities or training required.

In the 53 million gainful-worker group in 1940, nearly 12 million were women and over 41 million were men. For several decades the percentage of women in the labor supply has been increasing, partly because the development of service trades has increased opportunities for their employment. However, only about 22 per cent of all women in the age groups from 20 to 60 are in the gainful-worker class, and more than half of these are under 30 years of age. On the other hand from 90 to 98 per cent of the men in these age groups are gainful workers. Since virtually all men between 20 and 60 are in the gainful worker group, very little expansion in labor supply can be expected from this source at any given time. However, women who are not in the gainful-worker category represent a sizeable reserve of potential workers. This was illustrated during the war when a large part of this reserve entered gainful work, thus largely compensating for the loss of male workers to the armed services.

Long-Term Trends in Labor Supply

Changes in labor resources over a long period of time occur because (1) the total population changes and (2) the age composition of the population changes. The United States has experienced a phenomenal growth in population: from about 6 million in 1800 to 140 million in 1946. During the same period the labor supply grew from less than 2 million to about 60 million. The growth in population is accounted for by immigration and natural increase. Early in the nineteenth century birth rates in this country were high, about 40 per thousand population per year. Death rates were considerably lower and population therefore increased. Since then, birth rates have declined; in the early 1930's they dropped to 16 and in recent years have averaged about 20 per thousand population per year. Death rates declined too (to 11 per thousand in recent years), but the difference between the two rates has narrowed, consequently the rate of population growth has slowed. The drastic restriction of immigration since the early

1920's has also decreased the rate.[11] Population experts predict a balancing of birth and death rates in this country within a few decades. If this occurs the American population and eventually its labor supply will have reached maximum size unless immigration is permitted on a large scale. The British and French have a balanced rate at present; on the other hand, Russia and Asia generally are experiencing substantial increases.[12]

Age composition means the percentage of population within defined age groups. These groups are customarily defined as (1) under working age, or under 15 years, (2) working age, or from 15 to 64 years, and (3) over working age, or 65 and over. As a population increases or decreases, the percentage in each of these broad groups normally changes too. Thus in the early American period of growth, during the time of high birth rates, the percentage of those under 15 years was larger than it is today, since the birth rate has sharply declined. The percentage in the working-age group, 15 to 64 years, is larger now than a century ago. Thus a change in age composition may add to the labor supply over and above the additions caused by total population growth. For some time after a population becomes balanced or static, labor supply will grow because of the increase in the working-age group. But the progression of ages in a static population extends farther, so that the percentage of the working-age group will decline eventually as the over-working-age group increases. At this stage the number of people annually reaching 65 years of age will be larger than the number of young people attaining 15 years, and the labor resources of a country begin to decline.

It will be a number of decades at least before the cycle reaches this stage in the United States. Labor supply here is growing about 500,000 annually. But France with its gradually declining population, and Britain whose population is virtually at its peak, face a decline in labor supply before many years. Sir William Beveridge in his well-known Report said that Britain faces the possibility of World War III a generation hence with

[11] Between 2 and 5 million immigrants entered this country each decade from 1850 to 1900; from 1900 to 1909, 8 million entered. In 1913 and 1914 the annual average was 1.2 million. Over 40 million immigrants have entered this country since 1820. Beginning in the early 1920's Congress limited immigration severely, permitting since 1924 only 150,000 quota immigrants per annum. Immigration from Canada and Mexico is not subject to quota restrictions, but after 1930 all immigrants were required to prove that they would not become wards of public charity. These restrictions have slowed immigration and consequently population expansion.

[12] The decline in birth rates to 16 in the early 1930's has been reversed, especially during the war years. This rise in birth rates has added several million to the population above the estimates of future growth made in the prewar years. This fact alters the experts' predictions mentioned above, but probably will not affect their conclusion, namely, that population stability is probable in future decades.

several million fewer people in the working and fighting ages than it had in World War II.

The influences of population changes on a nation's productive capacity are quite evident. Growth in labor resources has been one of the foundations of the growth in production, here and across the Atlantic.[13] For several decades more at least this country is assured a continuing increase in this resource; at sometime in the future a stable, then perhaps contracting, labor supply is expected.[14] In the future, therefore, increases in productive capacity will depend more and more on the growth of capital and on the development of new energy and mineral resources.

GENERAL INFLUENCES GOVERNING THE USE OF RESOURCES

Economic resources—human effort, capital, and natural resources—exist in all economic systems. But the ways in which they are used and combined differ radically among countries. Production is organized economically so that wide use is made of the relatively abundant resources, limited and careful use of the relatively scarce resources. A combination of resources appropriate to India would be most uneconomical in the United States. Each country tends to use its resources in a way calculated to obtain the best productive results, i.e., the largest output at the least cost.

Production in different countries is also conditioned by the prevailing technology which is the knowledge and technique employed in using economic resources. Technology is not limited to mechanized techniques; handicraft work, for instance, is based on its own technology. Technology, in a broad sense, embraces the knowledge and techniques of organization, of marketing, of finance, and the whole range of operations through which economic activity passes. The characteristic technology of modern economic systems revolves about mass production, large-scale organization and finance, extensive marketing, and transport systems.

Social institutions—embracing customs, laws, generally accepted beliefs, legal concepts—profoundly condition the use of economic resources. In the United States law and public sentiment support the system of private property and private initiative, so that economic resources in this country are organized largely on the basis of these social institutions. Custom is

[13] Indeed, during the years 1890-1910, industry in general encouraged immigration as a means of acquiring a cheap labor supply for its expansion.

[14] Such changes may show variations; thus the increase in births during the 1940's over the 1930's will make labor supply additions in the 1960's greater than those of the 1950's. A change in one decade may be reversed in the following because of earlier variations in birth rate.

an important influence in governing economic wants—a producer ignores custom at great risk. Underlying beliefs, such as a conception of justice, hold economic activity within certain limits. These beliefs may be reinforced by law, such as those governing the employment of women and children, the length of the working day, minimum wages, the issuance of corporate securities, etc. In short, economic behavior, like all social behavior, is subject to and governed by sanctions, legal or customary.

PRODUCTIVE CAPACITY

Productive capacity is not an easily measurable limit. During a prolonged depression capacity might decline as labor loses its skill through idleness and capital facilities are allowed to deteriorate. A period of prosperity or of war, on the other hand, brings with it an expansion of capital. At any given time productive capacity depends upon (1) the volume and types of economic resources, (2) the allocation or distribution of these resources among the various uses for them, and (3) the skill with which these allocated resources are used.

These points need some amplification. The first relates to the quality, quantity, and "balance" or ratio among economic resources. Little more need be said of the importance of the quality and quantity of human resources, capital, and natural resources. But quantity and quality alone are not decisive if resources are not "balanced." India, for example, has an enormous labor supply, but its capital equipment and natural resources are inadequate; consequently its productive capacity per person is low. To be sure, the few have great wealth but most of the people are extremely poor because their great numbers mean low productivity and income. On the other hand, the balance or relative supply of resources in the United States is such that labor is amply supported by capital and natural resources; consequently productive capacity is high.

Nations sometime follow policies, such as imposing subsidies and tariffs, for the purpose of building up certain productive segments of the economy. An industrial nation, such as Germany, endeavors to develop agriculture and natural resources as fully as possible; and an agricultural nation, such as Brazil, endeavors to develop industrial resources. This balancing of resources is often sought for purposes of military security, e.g., a nation wishes to become independent of uncertain and distant sources of supply. The allocation of resources among the varied fields of production affects, too, the productive capacity of a nation. The value of product or output per worker is usually less in agriculture than in manufacturing, and less in

manufacturing than in the service industries. This may be explained by the fact that agriculture is usually an "old" industry with an ample labor supply, while manufacturing is newer and employs relatively less labor, as do the service industries. Predominantly agricultural countries have a relatively low productive capacity per worker; predominantly industrial countries have a relatively high productive capacity per worker.[15] Thus the same complement of economic resources will usually yield different outputs, and have different capacities, depending upon their allocation, i.e., degrees of scarcity, among the various productive fields.

The importance of intelligent and enterprising management as a factor in productive capacity is beyond question. Economic resources do not function automatically; they must be organized and directed. Management, either public or private, performs this function, and both actual output and productive capacity vary with the degree of skill and enterprise exhibited by management.

Long-Term Changes in Productive Capacity

Eventually, when population and labor supply decline, the United States and the western European industrial nations will no longer continue to grow in productive capacity. New industrial areas, such as Russia and Japan, with rapidly growing populations, will enjoy a relative advantage over western Europe and the United States so far as changes in labor supply are concerned. Unlike labor resources, there is no reason for capital growth to cease in the United States and Europe as long as (1) incentives to invest exist and (2) the public is willing to save or forego some current consumption, i.e., to allow an excess of production over consumption. Unless the growth of capital and the improvement of technology continue when labor supply becomes stable, productive capacity will level off; and when labor supply begins its decline, a loss in productive capacity can be avoided only if capital continues to increase; and at this point a further growth in capacity can be achieved only if capital expansion is accelerated.

Changes in productive capacity resulting from exhaustion or discovery of natural resources may show as great a geographical diversity in the future as in the past. It is not unlikely that the older industrial nations will experience considerable difficulties in maintaining their relative production positions because of natural resource depletion, while new sources of raw materials in Russia and other parts of Asia may favor expansion there. As new industrial centers, these latter regions have not expended their en-

[15] See Colin Clark, *The Conditions of Economic Progress*, Macmillan, 1940.

dowment of natural resources to the extent that Europe and the United States have. Productive capacity, therefore, may shift heavily in favor of Asiatic nations—a movement reinforced by their large and growing labor resources.

With all these influences affecting long-term changes in the productive capacity of a country, and its position vis-à-vis other productive areas, developments in technology play a major role. In many respects new technological developments assume a position of growing importance for the older industrial nations, with their stable or near-stable labor resources and diminishing reserves of metal and oil. The newer industrial areas, such as Russia, may be able to place relatively more dependence upon the elements that accounted for much of the growth of the older industrial areas, namely large quantities of man power and natural resources. This is no longer open to the United States and Europe to the same extent, consequently technological developments and capital investment must grow in importance as elements in future growth of productive capacity.

Production and the Capacity to Produce

The actual volume of production must be distinguished from the capacity to produce. There are times when the volume of production almost equals the capacity to produce, as it did during the war. At other times production is well below this limit, with excess or idle capacity existing in virtually all spheres of production, as during the prolonged depression of the 1930's. Given the productive resources or capacity, several different conditions under which this capacity is used may be distinguished—ranging from a maximum of a war economy to a minimum of a depression economy.

1. *Extended use of economic resources.* This is the condition during war periods, and involves setting aside many of the usual restrictions on the use of resources: increase of work per day and per week, curtailment of workers' time off, speeding-up of operations, abandonment of the single shift for two or three shifts in factories and mills, the deferment of normal repair work on capital facilities to reduce time lost, and a host of other expedients to get the maximum production out of the resources at hand.
2. *Relatively full employment of economic resources.* This is the condition during a period of normal prosperity. At such a time the customary limits to resource-use remain in force, and given these limits, virtually all labor willing to work at the wages and conditions offered

is employed. Virtually all, but not all, for at any time some workers will be idle because of illness and accidents or because of shifting from one job to another, unavoidable interruptions in production will temporarily throw some people out of work, seasonal layoffs will mean idleness for some, and the gradual decay of some industries will mean unemployment or underemployment of workers unable or unwilling to shift to newer fields. Unemployment for these reasons is not large—perhaps no more than 5 per cent of the gainful workers. Under normal circumstances this may be termed normal unemployment. For all practical purposes, this represents a condition of relatively full employment of the resources available for work. (The late 1920's and 1946-47 approximated this condition.)

3. *General unemployment of labor and idleness of capital facilities.* This is the condition during a period of depression. From 1930 to 1940 many millions of workers able and willing to work at the prevailing wages were unable to obtain employment. Depression unemployment was added to the normal unemployment discussed in the second condition. Capital facilities at the same time were idle on a large scale. The condition generally was one in which the volume of production was materially below the normal productive capacity of the nation.

The explanation of these changes in the level of production is one of the important tasks of economics, and one which is given considerable attention in the remaining chapters of this part.

SUMMARY

Production is a continuous process in which economic resources are brought together to produce commodities and services. This process may be viewed as a flow—an "inflow" of economic resources and an "outflow" of goods and services. Payments are made to owners of economic resources for their part in this process; these payments are the main costs of production, and as wages, etc., constitute the principal fund of purchasing power available to buy the output of the productive process. The size of this flow of goods and services depends upon (1) productive capacity, (2) the extent to which this capacity is used, and (3) the efficiency of this use. The allocation of these resources among the various fields of productive activity, their efficiency, organization, and management are elements which govern the capacity of the resources at hand. The extent to which this capacity is used in turn is governed by a complex of social customs and the monetary incentives which lead to decisions to produce.

Productive capacity and the flow of goods and services are subject to changes over time. In the past the opening up of new areas with their wealth of natural resources, a prodigious growth in population and labor supply, a substantial accumulation of capital, and development of technology all accounted for huge growth of capacity and output in the major industrial nations. In the course of time, some of these expansive influences are likely to recede in importance in the North Atlantic region. This may alter the rate of long-term expansion, and perhaps shift the center of growth to other areas, notably Asia. These long-term changes in productive capacity are only in part subject to control. The depletion of natural resources may be partly retarded by conservation in use, partly and perhaps wholly offset by the development of substitutes and new energy sources. The growth of capital is assured if incentives to invest exist and production is maintained at high levels, for capital is formed out of the excess of production over consumption. But little probably can be done to arrest the eventual decline in population and labor supply. The retarding effects of such a decline might be compensated by further growth in capital facilities, emphasizing the saving of labor, or increased productivity of labor.

As for the present and near future, the economic welfare and strength of a nation depend upon (1) the extent to which existing economic resources are employed, (2) the allocation of these resources among the various fields of productive activity, and (3) the distribution of the product, or income, among the population. The first of these matters is discussed in the remainder of this part, the second in Part IV, and the third in Part VIII. In the next several chapters the employment of economic resources to produce commodities and services is analyzed in terms of national income, its components, and the causes and conditions that govern its level.

CHAPTER 4

National Income and National Output

The commodities and services flowing from the productive activities of economic resources make up the national output, and the incomes earned in this process comprise the national income. This continuous flow of na-

tional output embraces a vast array of different commodities and services, of producer goods and consumer goods. The volume of this flow is of prime importance to the nation and to the individual because in general it measures the performance of the economic system. Since different units of measurement of commodities—bushels of wheat, bales of cotton, board feet of lumber, tons of steel, barrels of oil, gallons of gasoline, quarts of milk, yards of cloth—cannot be added together, the flow of national output is measured in terms of its dollar *value.* The total of commodities and services can thus be compared from year to year, therefore providing a means of gauging the performance of the productive system.

After World War I statisticians and economists became increasingly concerned with the problem of measuring the yearly value of the nation's output of commodities and services and its income. During the 1930's an acceleration of work on this problem resulted in the preparation of estimates by the United States Department of Commerce and the National Bureau of Economic Research of New York City.[1] (There are now annual estimates available beginning with 1914.)

Concurrently with the statistical measurement of national output and income in the early 1930's, economists made searching theoretical analyses to discover the causes that establish and change its level. John Maynard Keynes (later Lord Keynes) took the lead in this study.[2] The interest of economists and statisticians in national output and income was stimulated largely by the economic, political, and social problems which arose during the depression of the 1930's. The severity of these problems forced the attention also of political, business, and labor leaders to a consideration of the level of national income and output and the means of increasing it. When war broke out in 1941 the enormous military expenditures changed a prewar condition of unemployment of resources to a condition of extended use of resources with record levels of national output. With the close of the war economists and public leaders turned their attention to the prospects of maintaining high levels in the peace years following.

National output and income, therefore, have attained positions of special importance in economics and in public policy. Since the volume of national output generally measures the performance of the economic system, it is to a great degree a guide to public policy. Consequently some understand-

[1] *National Income* (Senate Document), 1932. A comprehensive revision of national income estimates appears in *National Income,* supplement to the *Survey of Current Business,* July, 1947, published by the United States Department of Commerce.

[2] His great work, *The General Theory of Employment, Interest and Money* (Harcourt, Brace, 1936) serves as the point of departure for much of the recent theoretical work in economics.

ing of national output and income is indispensable to the student if he is to penetrate even a short distance into the complexities of contemporary economic life.

MEANING OF NATIONAL INCOME AND NATIONAL OUTPUT

Thus far national output and national income have been defined only in general terms. It is necessary now to consider these concepts in more detail and to distinguish between closely related concepts and magnitudes. Specifically, five concepts are to be distinguished.

1. National income
2. Income payments (personal income)
3. Gross national output (or product)
4. Net national output (or product)
5. Aggregate demand or expenditure

These will be taken up, not in this order, but in a sequence which will bring out their relationships most clearly.

Gross National Output

Gross national output is the total output of commodities and services for a given period of time; this may also be termed *aggregate supply.* In 1946, for example, gross national output was estimated at 204 billion dollars. The commodities that make up this output are *finished* or *final products,* i.e., products in a form ready for final use. Consumer products, such as cigarettes, clothing, radios, food, are in a finished form ready for the final user—the consumer. Similarly, all currently produced capital facilities, such as machinery, equipment, and industrial buildings, are finished products ready for the final user—the producer.[3] War goods of all types are also finished products ready for the final user—the government. These three groups of finished products comprise the commodities that normally make up the greater part of the gross national output. *The value of these commodities is measured by their price (or cost) to the final user,* i.e., to the consumer, the purchaser of capital facilities, or the government.

In the production of finished commodities as defined here, raw materials, semifinished materials, and parts are processed, fabricated, and installed

[3] Capital facilities are finished or final products in the sense that they require no further processing or change; they are in the final form required for their intended use and in this form they normally are used over a period of time, i.e., they are durable-use producer goods.

into the final product. The value of the component elements used in the production of final products appears as part of the value of the final product. Thus, some part of the value or price of a loaf of bread "represents" the value of the wheat, some part of the value or price of an automobile represents the value of the steel used in its production. If the value of these materials were added to the value of the finished product, it is obvious some things would be counted more than once.[4]

The value of services forms the remainder of gross national output. All services, however, are not counted directly in measuring output. The services excluded are those necessary to commodity production; their value is "represented" as a cost in the value of the finished or final commodity. The services of all employees in a plant, for example, as well as the services of people engaged in related transportation, public utility, legal, and other occupations essential to the plant's operation are included in the value of its final products. But some services are rendered directly to individuals and, therefore, their value forms a separate part of gross national output. The services of beauty parlors, professional football and baseball players, physicians, lawyers, hotels; utility services for domestic purposes; services performed by government—all are in this category.

Gross national output therefore includes (1) finished or final products, such as consumer goods, capital facilities (both for replacement and net capital growth), and the materials and equipment for the armed services; and (2) all services of federal, state, and local governments, and services rendered by individuals and businesses directly to private persons. This total represents the results of the productive efforts of a nation's economic resources. In other terms, with the economic resources at hand—or a given productive capacity—the level of gross national output is a measure of the extent to which these resources are employed. Throughout this book when the term *national output* is used it will mean the gross output.

Net National Output

Net national output is equal to gross national output minus the depreciation of capital facilities. It is the net value of the commodities and services (as described above) produced during the course of a year, i.e., their value after allowing for the using up or depreciation of capital facilities during the year.

[4] For purposes of simplicity this discussion assumes that all raw and semifinished materials produced go into final products during the course of the year. Net changes in inventories are thus ignored at this point.

An example of a familiar type of small business enterprise may aid in making the distinction involved here. Suppose a man owns a moving van which cost him $4,000. During the course of a year suppose his gross receipts derived from moving family possessions come to $10,000. He pays $3,000 in wages to a helper, $1,300 for gasoline, oil, and repairs, and $200 in sales and other *indirect* taxes.[5] The $10,000 represents the gross value of the services rendered by this enterprise and is analogous to gross national output. During the year his van suffers wear and tear, and at the end of the year it is worth only $3,000. Depreciation, therefore, is $1,000. From the proceeds of his business he deposits $1,000 in his bank so that at the end of four years (assuming he deposits this amount each year) he will be able to replace his van. Therefore, the $1,000 of depreciation is deducted from his gross receipts, leaving $9,000 as the net value of the services rendered during the year by this business; this is analogous to net national output. The $1,000 wear and tear is deducted because it represents that part of the value of the van used up during the year. Or it might be said that this is the amount of last year's output (when the truck was built) that appears in the gross value of this year's services rendered by the enterprise.

A similar illustration may be employed for the economy as a whole. Suppose this nation produces in some one year a gross national output of 190 billion dollars. The capital facilities used in the production of this output will suffer wear and tear in the process and become one year older during the course of the year. For both reasons, the value of the capital facilities will be less at the end of the year than at the beginning: this loss is *depreciation.* The capital thus used up is in effect converted into finished commodities in the process of production. Suppose this depreciation amounts to 13 billion dollars during the year. In order to restore these capital facilities—or to maintain them—13 billion dollars of the gross output (i.e., in the form of new capital facilities) must be devoted to replacement of these facilities. In the previous chapter this was referred to as *capital replacement.* When this replacement is accounted for the balance is 177 billion dollars, the *net national output.* The value of net national output is equal to all the incomes *earned* or *produced* in the course of the year plus indirect taxes levied on business which appear in the cost of commodities and services produced.[6] To continue the illustration it is

[5] The indirect taxes paid by this entrepreneur are part of the cost of the services rendered by him and must be distinguished from a *direct* tax, such as the personal income tax, which is not a business cost.

[6] Department of Commerce estimates include other items as well which may be ignored for present purposes.

assumed that indirect taxes are 10 billion dollars and incomes 167 billion dollars.

National Income

National income is the *total of all incomes earned in the productive process* during a given period of time. These incomes are the *factor costs* of the commodities and services produced, and take the form principally of wages, salaries, profits, interest, dividends, and rent earned in production. National income is equal to net national output minus (principally) indirect business taxes; or, factor costs plus indirect business taxes equal net national output.

The national-income concept may be illustrated by further examination of the moving-van example. It was shown that the owner of the van set aside $1,000 for depreciation, and paid indirect business taxes (e.g., the sales tax on gasoline purchased for the van) amounting to $200. When this tax is deducted from the *net* value of services rendered ($9,000) the $8,800 remaining is the income earned (directly or indirectly) by the operation of the van during the year. The owner paid his helper $3,000 in wages; and he spent $1,300 on supplies and repairs. This amount indirectly is income for it may be resolved into incomes of those who sold him the supplies or rendered him services, e.g., the mechanic who repairs the van. The amount remaining—$4,500—is the income which accrues to the owner of the van. These incomes total $8,800 and this total is analogous to national income.

The distinctions between gross output, net national output, and national income can be made clear by developing the previous illustration using aggregate figures. The production of commodities and services involves costs. What are the costs of a gross national output valued at 190 billion dollars? First, the cost of the capital facilities used up in producing the output of commodities and services (13 billion dollars); second, the indirect business taxes (10 billion dollars) which are included in the prices of finished products to their buyers; and third factor costs, the wages, salaries, interest, rent, profits, and dividends involved in production (167 billion dollars). The total cost of producing gross national output therefore is 190 billion dollars; that part accruing to the factors of production, 167 billion dollars, is national income, and this, plus indirect business taxes, is net output. The 13 billion dollars of capital used up is attributable to efforts of earlier years when the capital was produced; the 167-billion-dollar income measures the net current effort required for production.

Income Payments (Personal Income)

A distinction must be made between income *earned* or produced in the economic system and the actual income *payments* flowing to the public at large. At first sight, it might appear that these should be identical, and under some circumstances they might indeed be equal. It is conceivable that the 167 billion dollars of national income in the present illustration would be accompanied by income payments (wages, salaries, profits, etc.) of the same amount. While possible, this is not usual, because (1) some earned income is not paid out to those who have a claim to it; (2) at other times more income is paid out to those participating in production than is actually earned; and (3) at all times income payments are made to some groups whose claim to these payments is not based on current productive effort. Thus in practice income payments usually exceed or fall short of national income; only rarely are they equal.

An *excess* of income payments over national income usually occurs during a period of depressed economic activity. This excess occurs because:

1. Government makes substantial payments to groups who do not produce goods or render services in return, e.g., relief payments to the unemployed and the destitute aged, bonuses and benefits to veterans, etc.; [7] and it pays large amounts as interest on its debt.
2. Business enterprises, particularly corporations, make income payments—such as dividends—although this income may not be currently earned. Funds from a company's bank balance might be drawn on for this purpose, or, more likely, part of the current receipts are used to pay dividends instead of being used to replace capital facilities.

Part of the excess of income payments over national income in the depression years of the 1930's was due to the attempt on the part of government to stimulate economic activity by spending public funds. More will be said on this point later.

Income payments fall short of national income as a rule in a period of normal prosperity and during war. The reasons for this deficiency of income payments may be summarized as follows:

1. Business retains part of its profits, holding them as "undistributed earnings," when earnings are large. This occurred during the prosperous 1920's and again during and immediately after World War II.

[7] Such payments to "nonproducers" are often called transfer payments.

2. Part of corporate income is taxed under corporate income-tax laws and therefore is not available to be paid by the corporations to their owners; also social security taxes are treated as income but are paid to government, not to those who earn the incomes.

The distinction between income earned and income payments may be brought out by the moving-van illustration. Income earned is $8,800; the wage payment is $3,000, and $1,300 is resolved into other income payments, leaving $4,500 as the earned income of the entrepreneur. If he draws this out, income payments and income earned are equal. He may, however, take only $3,500 of his earnings and add $1,000 to his business; in this case income payments of $7,800 are less than income earned ($8,800). On the other hand, suppose he draws out the $4,500 as well as the $1,000 previously set aside for depreciation. In this case his income payments (to himself) are $5,500, and income payments exceed income earned. Business often pays out income during depression by failing to set aside reserves for depreciation; and in good years it normally pays less than it earns; i.e., it saves.

Aggregate Demand or Expenditure

Aggregate demand is composed of (1) the expenditures of consumers for commodities and services, (2) the expenditures of business for capital facilities and other investment goods, both for replacement purposes and for capital expansion, and (3) the expenditures of government for commodities and services. These expenditures are equal to the value of gross national output; the value (or the sum of the market prices of commodities and services comprising gross national output) is measured by aggregate demand—indeed it is determined by aggregate demand.

Gross Output, Net Output, National Income, and Income Payments

In the discussion above these basic concepts have been briefly described and contrasted. Table 2 gives figures for each year since 1929. Column 1 is the gross national output, which is the dollar value of all final commodities produced, services rendered to persons, and the cost of governmental services. It measures the over-all productive activity or performance of the economic system. Column 2 is net national output. The difference between columns 1 and 2 measures chiefly the amount of capital used up in producing gross output. Column 3 is national income, total income

earned each year, and column 4 is the payment of income which comprises current purchasing power in the hands of the public. This latter amount is available for consumption expenditures, individual savings, and the payment of personal taxes.[8]

Table 2 NATIONAL OUTPUT AND NATIONAL INCOME, 1929–46

	1. Gross national output	2. Net national output	3. National income	4. Income payments
1929	$104 billions	$95 billions	$87 billions	$85 billions
1930	91	82	75	76
1931	76	68	59	65
1932	58	51	42	49
1933	56	49	40	47
1934	65	58	49	53
1935	72	65	57	60
1936	85	77	67	71
1937	90	82	74	74
1938	85	77	67	68
1939	90	82	73	73
1940	100	92	81	78
1941	125	116	104	95
1942	160	150	136	122
1943	193	182	168	149
1944	210	199	182	165
1945	213	201	183	172
1946	204	193	178	177

Source: U.S. Department of Commerce, *National Income*, Supplement to *Survey of Current Business*, July, 1947.

Gross national output measures the "real" output of the nation in the course of a year; it comprises the commodities and services for the satisfaction of consumer wants, the currently produced capital facilities for the productive system, and the materials and equipment of war. This production activity is the chief source of the nation's income payments; and these payments are the principal source of purchasing power to buy the output of this productive activity. Except for that part produced by government, gross national output is produced for sale principally to consumers and

[8] Current expenditures of individuals are of course supplemented by expenditures from "liquid assets" such as bank accounts and government bonds. Thus purchasing power is not limited to current income.

producers. As this output flows to the final users or purchasers, most of the income paid out in producing it is returned to producers in the form of consumer expenditure; the balance of output is purchased by the other components of aggregate demand—investment and government expenditures. There is, then, a great circular flow of output, money incomes and other costs—an *outflow* of commodities and services and a *return flow* of money as these commodities and services are purchased.

COMPOSITION OF NATIONAL OUTPUT

Commodity and Service Production Volumes

Commodity production in this country normally accounts for approximately two thirds of the value of gross national output. Services other than those essential to the production of commodities (transportation, public utilities, financial services, etc.) are those rendered directly to individuals (medical, legal, educational, recreational services, etc.) and the services performed by government. The latter are measured by the incomes paid to governmental employees. Together, these services account for the remaining one third of gross national output.

The quantitative importance of the "service industries" is considerably greater than might be inferred from the fact that the value of services accounts for but one third the gross national output because a substantial part of the value of final commodities represents the value of services essential to their production. The dominant importance of service production is a reflection of the highly developed economic life of this country. A simple economy depends largely on food production, construction, and a limited volume of manufacture, without much need for extensive transportation and public-utility systems, highly developed marketing mechanism, banking and insurance, or extensive service trades—hotels, service stations, beauty shops, restaurants, etc. But in the United States service industries employ about as many workers as do commodity-producing industries, such as agriculture, manufacturing, mining, and construction. As a nation grows in industrial wealth the importance of services in the gross national output increases.

Consumption and Nonconsumption Output

The composition of national output may also be examined in terms of the purposes served by commodities and services. These include:

1. Consumption
   a. Commodities
   b. Privately produced services
   c. Governmental services to consumers
2. Nonconsumption
   a. Capital facilities
   b. Governmental services to business enterprise
   c. War materials and equipment

Consumer goods and services (*consumption*) during the period 1929-40 comprised between 70 and 75 per cent of gross output. During the war, gross output increased greatly, principally in the production of war goods, so that consumer goods and services accounted for only 52 per cent of total output in 1943-44.

Capital facilities as a component of gross national output are subject to considerable variations. In periods of relatively full employment nearly 20 per cent of total output is in capital facilities; this includes both replacement capital and net investment.[9] In depression years the output of capital facilities is low; less than 10 per cent of total output may be in this form. At this depressed level the output of capital facilities may not be enough to account for full replacement of obsolete capital facilities, and the nation's productive capacity declines accordingly.

Governmental services as a component of national output vary in importance, largely because total output itself is subject to large changes. In 1929, these services accounted for 10 per cent of the total, in 1932 for 20 per cent, in 1940 nearly 15 per cent. Although the aggregate volume of governmental services has tended to increase steadily over the years, their importance as a component in general varies inversely with national output. In total dollar amount governmental output (services) varies less than any other component of the national output. Since income payments originate largely in the productive process, it follows that income payments from governmental sources are relatively constant (compared with other sources of income payments) from year to year.

War materials and equipment as a component of gross national output are subject to the greatest variations. Before the war, they accounted for a negligible fraction of total output; for example, 2 per cent in 1939. By 1943-44, war materials and equipment comprised over 40 per cent of gross output, which itself was more than doubled over prewar levels.

[9] As stated in the previous chapter, capital facilities comprise the bulk of the investment goods category; they deserve emphasis, too, as a basic element in the nation's capacity to produce.

Thus the components of gross national output—both for consumption and nonconsumption purposes—vary considerably with peace and war, and with prosperity and depression. To repeat, these components represent the "physical" output of the nation's economic resources, and their production creates or initiates the greater part of national income payments which constitute the major source of purchasing power to buy this output.[10] The economic experience of this country shows that when the output of capital facilities (and other investment goods) is large, the economy is relatively prosperous; when low, the economy is depressed. *Except during war, the decisive influence governing the level of national output and income payments is the volume of investment goods production, especially capital facilities.*

Industrial Sources of National Output

The industrial sources of gross national output—farms, factories, mills, mines, etc.—reveal the importance of the various productive segments and therefore the general character of the economy. Table 3 shows the percentages of the 1941 national output produced by the major productive groups in this country.

Table 3 **SOURCES OF NATIONAL OUTPUT, 1941**

Sources	Percentage
NATIONAL OUTPUT	100.00
Manufacturing	28.9
Trade (retail, wholesale)	15.1
Government (federal, state, local)	12.0
Service (legal, medical, hotel, service stations, etc.)	8.4
Finance (banking, insurance, brokerage, etc.)	8.0
Agriculture	7.7
Transportation (rail, water, truck, air)	6.8
Contract construction	3.6
Mining	2.2
Power and gas	1.9
Communication (radio, telephone, telegraph)	1.1
Miscellaneous commodities and services	4.3

Source: U.S. Department of Commerce.

[10] This is true except during a war economy such as in the United States in 1943-44 when government purchases nearly equaled private purchases of the gross output.

Manufacturing, trade, and government, in that order, are the largest contributors to national output and to national income payments. The importance of each of these sources of income, however, changes as underlying economic and political changes occur. Agriculture once accounted for the bulk of national output; but the growth of manufacturing and the collateral expansion of transportation, trade, and the other service industries has relegated agriculture to a subordinate position within the space of a century. Long-term changes in the relative positions of the main productive sources therefore reflect underlying structural changes in the economy.

In the short term other changes occur because of the rise and fall of national output, because of war conditions, or the transition from a war to a peace economy. For example, manufacturing accounted for nearly 29 per cent of national output in the prosperous year of 1941, but for only 15 per cent in the depressed year of 1932. In 1945 manufacturing accounted for about 50 per cent of national output, a large part of which was devoted to war goods. At the close of the war, the reduction in war-goods production brought about a decline in the importance of manufacturing as a source of national output. In a highly industrialized economy the principal changes in the levels of national output occur in the manufacturing sector, and particularly in the production of capital facilities and durable consumer goods.

These industrial sources of national output and national income payments are interrelated in a complex way. The amount of commodities manufactured obviously affects the volume of transportation, trade, the amount of minerals drawn from the earth, etc. At the same time, the volume of output and income payments in the nonmanufacturing components has an important influence on the level of manufacturing activity. For example, when manufacturing is prosperous, railroads are busy transporting large quantities of manufactured commodities, and thus require more locomotives, freight cars, and other capital facilities produced by the manufacturing industries. Further, the large volume of wages and other incomes paid out by the railroads adds to the general demand of the public for the consumer goods produced by the manufacturers. Each of these productive sources, then, is dependent upon the others, and each one in turn affects the level of activity of the others.

DISTRIBUTION AND USE OF NATIONAL INCOME

It was shown above that the value of gross national output is equal to (1) total income currently earned in the production process (national in-

come), (2) the costs to producers for capital replacement, and (3) indirect business taxes. Total income earned—the national income—represents the factor costs of producing national output. National income is divided up among those who participate in the productive process in the form of *distributive shares.*

These shares go principally to those (1) who work for wages, salaries, and the expectation of profits, and (2) who have claims to income based on ownership of property or assets. Table 4 gives the distributive shares for 1941—a year that reflects neither the depressed conditions of the pre-war period nor the extended use of resources characteristic of the later war years.

Table 4 **NATIONAL INCOME BY DISTRIBUTIVE SHARES, 1941**

Compensation of employees (mostly wages and salaries)	$64.2 billions
Income of unincorporated enterprise: business and professional, 9.5; farm income, 7.0	16.5
Rental income of persons	4.3
Corporate profits (before corporate income tax) (including inventory valuation adjustment)	14.6
Net interest	4.1
Total	**103.7**

Source: U.S. Department of Commerce.

This table represents income *earned* during the course of the year by those who participate in the production process. As stated above, income earned is not the same as the income payments people actually receive. For example, the figure showing compensation of employees includes taxes paid for social insurance and contributions of employers to private pension funds for their employees; these are regarded as employee earnings held for them by the public and private insurance systems. Also the corporate profits earned are by no means all paid to stockholders; in 1941 the corporate income tax receipts amounted to nearly 8 billion dollars, and undistributed earnings to nearly 5 billion dollars; of the 14.6 billion dollars of corporate profits, dividends amounted to only 4.5 billion dollars.

Nearly two thirds of the total is labor income—wage and salary compensation; this share is usually fairly constant. The shares that show the greatest percentage changes are net incomes of corporations and net income of proprietors. These drop sharply in depression years; indeed, during the worst years of the 1930's, corporate net income was a minus figure, i.e.,

corporations in general suffered net losses. The percentage of interest payments, on the other hand, increases during a depression period; in 1932, for example, it was 14 per cent, compared with 5.2 per cent in 1941. The cause of this is that aggregate interest payments for the prewar decade ranged from 5 to 6 billion dollars a year; therefore in years of depressed national income the percentage of this item to the total was relatively large, but in years of high national income it was relatively low.

The Disposition of Income Payments

Each year the public at large has at its disposal money in the form of income payments, received as "productive income" and as transfer payments. These income payments over the last decade or so ranged from 47 billion in 1933 to 177 billion dollars in 1946. In Table 2, on page 92, the estimates are given for each year beginning with 1929. These income payments represent a financial flow accompanying the vast stream of goods and services making up the national output and comprise the principal costs of this output.[11] The rise and fall of this money-income flow depends almost entirely upon the flow of physical product.

For what purposes are these incomes used?

Normally, the great bulk of annual income payments is spent for consumption, or the direct satisfaction of wants—food, clothing, and shelter, the "comforts and conveniences" of life. In an earlier section of this chapter it was shown that normally the greater part of national output takes the form of consumption goods and services. Consumer expenditures for the immediate satisfaction of wants directs this productive activity. Another part of income goes into savings, in the form of deposits in savings banks, building and loan associations, and other banking establishments; premiums paid to life insurance companies; purchases of stocks and bonds; payment of debts; and purchases of capital facilities. Savings may be defined as that part of income not used for immediate consumption or for the payment of personal income taxes. A third part of income is paid out by individuals as *personal income taxes.*

The proportions accounted for by consumption, saving, and personal income taxes are subject to considerable change over a period of time. One factor affecting these proportions is the *level* of total income payments. Normally the percentage of total income saved increases as total income

[11] These income-payment figures are not entirely money payments to individuals. Actually, the estimates include some incomes in kind, and some payments to institutions; however, for present purposes these qualifications are unimportant.

increases; the population tends to save a larger percentage of a 150-billion-dollar national income than it does of a 75-billion-dollar income. Another factor is the personal distribution of income, i.e., the way income is distributed among the individuals in a population. The less unequally income is distributed, the lower the volume of savings.

Table 5 shows total income payments for 1941 and 1946 by amounts accounted for by consumption expenditures, savings, and personal income taxes. It will be noted that the term *disposable income* is applied only to that part of income remaining after personal taxes are paid. Tax payment is a matter of compulsion; but once the claims of the tax collector are satisfied the balance of income is available for personal disposal, for consumption expenditures and savings.

Table 5 DISPOSITION OF INCOME

Income payments	*1941*	*1946*
Income payments	$95.3 billions	$177.2 billions
Less: personal taxes	3.3	18.8
Equals: disposable income	92.0	158.4
Less: consumer expenditures	82.2	143.7
Equals: personal savings	9.8	14.7

Source: National Income, Supplement to *Survey of Current Business,* July, 1947. U.S. Department of Commerce.

Consumption is by far the dominant use made of income. However, as income increases the percentage of income payments spent for consumption tends to decrease; on the other hand, the percentage saved tends to increase as income payments rise. In Table 6 the percentages of income

Table 6 PERCENTAGE DISPOSITION OF INCOME PAYMENTS

	Income payments	*Consumption expenditures*	*Personal savings*	*Personal taxes*
*1933**	$47 billions	99.3	2.2	2.9
1939	73	93.0	3.7	3.3
1929	85	93.0	4.1	2.9
1941	95	86.4	10.2	4.4
1943	149	68.0	20.0	12.0
1946	177	81.0	8.3	10.7

Source: National Income, Supplement to *Survey of Current Business,* July, 1947. U.S. Department of Commerce.

*The percentages for 1933 add up to 104.4 because past savings were drawn upon in this year to supplement income payments.

payments spent, saved, and paid in personal taxes are shown for a number of years, which are arranged in order of the magnitude of total income.

The relationship between income payments and consumption expenditure is of utmost importance; it is termed the *propensity to consume* or the *spending* (or consumption) *ratio*. Thus, the spending ratio normally is lower for a high income, and higher for a low income. Of the 47 billion dollars of income payments in 1933, 99 per cent was spent; in 1941, 86 per cent of the 95-billion-dollar income was spent; and in 1946, 81 per cent of 177 billion dollars was devoted to consumption. This relationship is not unvarying; the spending ratio was higher in 1946 than in 1943, although income in 1946 was higher. In 1943 consumer expenditures were abnormally low (compared with income) because of wartime controls and restraints. Many of these had disappeared in 1946 and consumer expenditures rose.

The relationship between income and savings is termed the *propensity to save* or the *savings ratio*. This relationship is such that the percentage saved normally increases as income rises, and declines as income falls.

There are exceptions to the pattern of spending and saving as given here. The 1943 and 1946 changes in spending ratios have already been noted. War and postwar conditions also affected savings ratios. During the war years the pressure to save was insistent; government encouraged savings as a means of curbing consumer expenditure. Savings in 1943 (and in the other war years) were unusually high, comprising one fifth of income payments. In 1946 savings declined sharply, partly because consumers sought to buy many of the things not available during the war years. Moreover high income-tax rates in 1946 probably reduced the savings ratio.

The percentage of income payments absorbed by personal taxes for the years 1929-45 shows a steady increase, from less than 3 per cent in 1929 to more than 10 per cent in recent years. Depression and war have increased the activities of government, and these activities drain off a growing proportion of income into the public treasury. As noted above the rise in taxes affects the saving (and undoubtedly the spending) ratio.

The Composition of Output and Disposition of Income

To repeat, the value of gross national output equals the value of consumer commodities and services, capital facilities and other investment goods, and governmental services. And the financial counterpart of this output equals income payments plus business savings (undistributed profits) plus

business taxes plus funds retained by enterprise for depreciation purposes. The financial counterpart, then, is equal to the value of gross national output.

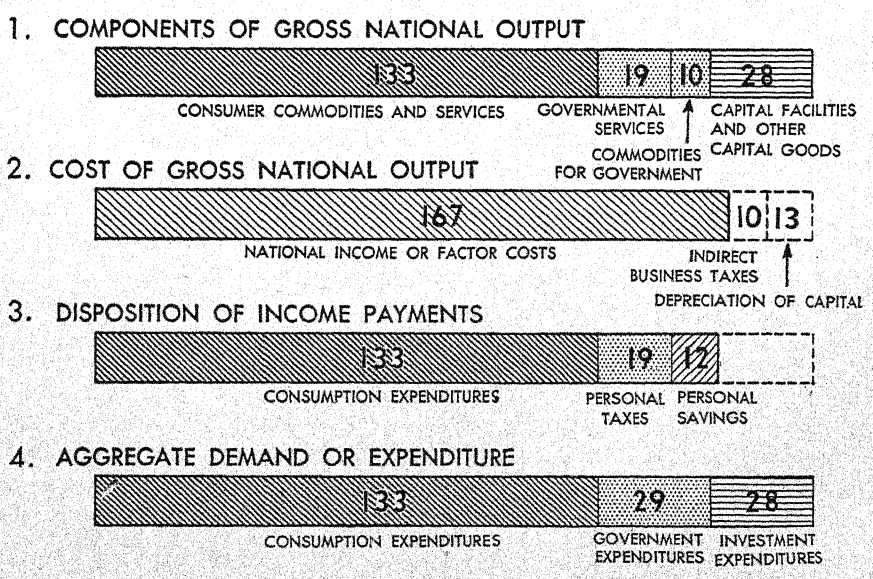
The foregoing is shown in Chart 1, based on the assumption of a gross national output of 190 billion dollars, which is taken to represent a condition of relatively full employment. The top bar on the chart (bar I) represents the value of gross national output shown as (1) consumer commodities and services, (2) services produced by government (value equal to the wages and salaries paid government employees), (3) privately produced goods for governmental purposes (e.g., war goods), and (4) capital facilities and other investment goods. As already mentioned these commodities and services may also be termed aggregate supply.

The second bar represents the costs of producing the gross national output: (1) national income or factor costs, 167 billion dollars, (2) indirect business taxes, 10 billion dollars, and (3) depreciation of capital, 13 billion dollars (which may be assumed to be represented by reserves set aside for depreciation).

The third bar shows the disposition of income payments—consumption expenditures, personal savings, and personal taxes. Income payments are

Chart 1 **GROSS NATIONAL OUTPUT, NATIONAL INCOME, INCOME PAYMENTS, AND AGGREGATE DEMAND**

FIGURES IN BILLIONS OF DOLLARS



164 billion dollars (and undistributed profits of 3 billion dollars make up national income). Business reserves for depreciation, business undistributed profits, and indirect business taxes comprise the remainder of this bar.

The fourth bar is aggregate demand or expenditure made up of (1) consumer expenditures of 133 billion dollars, (2) governmental expenditures of 29 billion dollars (equal to personal taxes and indirect business taxes), and (3) investment expenditures of business (equal to personal savings, business undistributed profits, and depreciation reserves) amounting to 28 billion dollars.

SOME STATISTICAL LIMITATIONS

Real and Current Income

The figures for both national output and national income given in the tables in this chapter are measured in "current dollars" for each year. For example, the value of commodities produced and services rendered in 1940 is measured by the prices prevailing in 1940, and so it is with the value of national output for each of the years shown. The total goods and services produced in 1940, multiplied by this average price over the year, is the dollar value of the national output for 1940.

Since both prices and quantities of goods and services together make up the total, it is evident that changes in either prices or quantities, or both, affect the total.

Gross national output varied from 104 billion dollars in 1929, to 56 billion dollars in 1933, to 213 billion dollars in 1945. Approximately half of the decline of 50 per cent from 1929 to 1932 may be attributed to the fall in prices during these years. The actual decline in the physical output of goods and services amounted to only about 25 per cent. Thus if these goods and services are valued in terms of, say, 1929 prices the decline from 1929 to 1932 would be 25 per cent, not 50 per cent.

Similarly, price increases during the war and postwar years account for part of the enormous rise in national income and output in this period. From 1939 to 1942 the value of national output measured in the prices prevailing in each of these years increased 70 per cent, but the volume of goods and services, when measured in "constant" prices, increased 50 per cent. That is, if the output of 1940, 1941, and 1942 is measured in terms of 1939 prices (calculated as though no change had occurred in prices after 1939) the influence of the price increases of those years is eliminated. The figures in Table 7 show this in terms of gross national output. Column 1

reflects changes in physical output and prices; column 2 shows price changes and 3 shows totals "corrected" for price increases. Such a correction measures changes in *real output.*

Table 7 CURRENT AND REAL OUTPUT, 1939–42

	Gross national output measured in current prices	*Price index*	*Gross national output calculated in 1939 prices*
1939	$ 90 billions	100	$ 90 billions
1940	100	101	99
1941	125	106	118
1942	160	114	140

Source: U.S. Department of Commerce.

For some purposes it is necessary to eliminate the changes caused by price variations. The well-being of the nation and the performance of economic resources from year to year are best measured by estimates of real national output.

Limitations of National Output Measurement

National output statistics of necessity underestimate the productivity of the economy. For example, the commodities and services produced by casual labor are not included because there is no way of measuring or estimating the total. Furthermore, the services of housewives are unpaid and cannot be measured or estimated in terms of money, yet they comprise an important part of want-satisfying activities. Actually, then, national output totals are restricted to (1) those commodities and services which enter the markets of the country in exchange for money, and (2) some goods and services which do not enter the markets but which can be measured by the prices of comparable marketable goods and services.

SUMMARY

Productive output is a measure of the performance of the national economy and has been described in terms of:

1. Gross national output, or the value of commodities and services produced.
2. Net national output, equal to the gross, minus the capital facilities used up in production.

3. National income or factor costs, equal to net output minus indirect business taxes.
4. Income payments.

Gross national output is the outcome of the productive work of the nation's economic resources. In any year—say Year One—production is carried on with a labor force, working with capital facilities and natural resources, under the direction of the nation's managerial talent. At the beginning of Year One, suppose the capital facilities of agriculture, manufacturing, transportation, and all other productive segments were valued at 160 billion dollars, and 60 million gainful workers were at work producing goods and services.

In the course of Year One a stream of commodities and services is produced aggregating 190 billion dollars for the year. This output is the *gross national output*, representing the value of all *finished* commodities—consumer, capital, and war commodities—and the services rendered directly to consumers and other final users. The value of this output represents (1) the value of the services currently rendered by economic resources, measured by the wages, salaries, interest, profits, etc., paid to them, (2) the value of existing capital facilities used up in producing this output, i.e., the depreciation of these capital facilities, and (3) indirect business taxes.

Suppose that the 190-billion-dollar gross national output of Year One represents 167 billion dollars of current services of economic resources, 13 billion dollars of capital facilities used up, and 10 billion dollars of indirect business taxes. If the capital facilities of the country are to be maintained at their 160-billion-dollar level, it is clear that 13 billion dollars of current output must take the form of capital goods and be used for capital facility replacements. This merely replaces capital facilities built in previous years and used up in Year One. In order to measure the net output of the productive activities of Year One, the 13 billion dollars is deducted. This is *net national output* which measures the productive performance for Year One without including the "help" derived from previous years' capital accumulations.

If the capital goods produced in Year One amount to 13 billion dollars and no more and the remaining 177 billion dollars represents consumer commodities and services and government services, then Year Two begins with the 160-billion-dollar capital facilities intact, neither more nor less than Year One began with.

But suppose 28 billion dollars of capital facilities and 162 billion dollars

of consumer goods and governmental services were produced in Year One (see Chart 1). This means that in addition to replacing wornout capital facilities 15 billion dollars (*net investment*) is added to the 160 billion dollars of facilities.[12] Thus Year Two would begin with 175 billion dollars of capital facilities which would enable the economy (under favorable circumstances) to increase both gross and net output in Year Two.

However, if in Year One only 3 billion dollars of capital facilities were produced, capital facilities used up during the year would not be fully replaced. Year Two would begin with total capital facilities down to 150 billion dollars, which means an impairment of productive capacity and probably a reduction in gross national output in Year Two.

National income payments are made to those who contribute time, effort, and assets in the productive process, in addition to relatively small amounts to those unable to contribute currently to productive activity. In Table 2 on page 92 income payments are shown together with estimates of output and national income. For every year shown, income payments were considerably less than the value of gross national output. This is so because business retains from its sales receipts the funds to provide for the replacement of worn-out capital facilities and for the payment of indirect business taxes.

The gross national output represents the supply of things available while income payments represent a large part of the potential demand for these things. Output is largely produced in expectation of profit and this expectation is satisfied when total output is purchased by the aggregate demand of consumers, investors, and government at prices expected by producers. If these purchases fall short of expectations, either because consumption expenditures decline, or purchases of capital facilities and war goods drop off, the *flow-back* of funds to producers is less than the *outflow* of income payments and other costs. Under these circumstances production declines, for producers get back less than their costs.[13] On the other hand, when aggregate demand exceeds income payments and other costs, producers get back more than they expected (i.e., their selling prices or sales volumes are higher than expected) and gross national output increases.

The extent to which economic resources are used in general depends upon the amount of consumption, capital, and war goods which producers expect to be able to sell profitably. The volume of consumption expendi-

[12] Net investment in this case excludes inventory changes and other forms of investment.

[13] Their receipts might cover actual money costs but not be sufficient to yield a profit; in this case output would probably decline.

tures depends upon consuming habits and the amount of income the public receives; the volume of investment expenditures depends upon prospective profits to be derived from private investment and upon the amount of governmental investments being carried out; the volume of expenditures for war goods, and for other governmental purposes, depends upon conceptions of national security and upon the purposes of government. At times these combined expenditures, or aggregate demand, are such as to tax the capacity of economic resources to meet the demands; at other times the combined expenditures are low, far below the capacity of existing economic resources to meet the demands.

Because these changes in gross national output are of utmost importance in governing the standards of living of the nation and defining its position in world affairs, the discussion will now proceed to the factors affecting the level of national output.

CHAPTER 5

Consumption, Savings, and Investment

The national output is a result of a complicated series of economic decisions—made by producers and consumers, by private industry and public authority. Producers decide to invest in buildings, machines, equipment, and other capital facilities, to buy materials, to hire labor, and to spend funds in effecting these decisions. It is these actions that initiate the flow of goods and services—and the money income—for the satisfaction of wants. To understand the operation of the economic system it is necessary to know why such decisions are made in the first place, and what circumstances influence them. The preceding chapter showed that most of the national output is produced by private enterprisers, while the relatively small remainder is produced by government. The decisions made by these two types of producers together—what may be termed *private* and *public decisions*—initiate total national output and national income payments.

Profit making is the primary purpose underlying the productive activities

of private enterprise. Under what circumstances are producer decisions made? What are the conditions that give reasonable assurance of profit? One condition, stated very generally, is obviously essential: things produced must be sold at profitable prices. Private decisions to produce are therefore guided by the general character of the *demand* for the goods and services they bring into production. This demand is expressed by the expenditure of money in purchasing these goods and services from three main sources: the consuming public (finished goods and services), private enterprise (capital facilities and other producer goods), government (both consumer and producer goods in small amount, except for wartime conditions).

Private decisions to produce, which account for 85 to 90 per cent of the gross national output, are therefore made in relation to demand for goods and services, or more correctly, the *expected demand.* Demand is measured by the amounts of a good which would be purchased at given prices; expected demand consists of the amounts producers expect buyers to purchase at possible prices in the future. Farmers and manufacturers, for example, begin the production of things today which they expect to sell many months hence. Farm machinery is purchased this year to produce crops for the next 10 years; a power plant is built this year to produce electricity for the next 50 years. The demand today for food, electric power, and automobiles was anticipated months, years, even decades ago when the capital facilities needed for their production were installed.

For purposes of the public interest, many economic decisions are made by government. A general "demand" also guides public decisions. This demand is not expressed by the purchase of government services.[1] Public health, public education, protection of person and property—these meet a public demand although they are not sold to the public. The need for them, therefore, is not expressed by a willingness to buy, as is the case with virtually all privately produced goods and services; it is expressed politically through legislative and other channels of government.

For purposes of analysis the present and future demand of the consumer public, private enterprise, and government has been termed the *aggregate demand* for national output; and the national output the *aggregate supply.* In the sections immediately following the profit expectations that lead producers to contribute to the aggregate supply shall be examined. This will be followed by consideration of the aggregate demand for national output, showing the parts played by (1) the demands of con-

[1] This statement holds for most governmental services and products. However, a publicly owned utility system operates in virtually the same way as private enterprise and is subject to the guidance of demand as expressed by purchases.

sumers in general for consumer goods and services, (2) the demands of productive enterprise for capital facilities and other investment goods, and (3) the demands of government for public purposes.

As discussed in the previous chapter, the cost of producing (or the value of) the gross national output approximates total income payments, business savings, the depreciation reserves of business, and indirect business taxes. Business savings and depreciation reserves may be termed *business gross savings.* Therefore, if the aggregate supply is to be sold at its value (or cost) aggregate demand must equal total income payments plus business savings and depreciation reserves plus the amounts collected by government as indirect business taxes. For purposes of illustration assume the value of national output to be 100 units; then aggregate demand for this output must be 100 if all costs are to be covered. If aggregate demand is less, some producers will suffer losses and production accordingly will fall; if aggregate demand is more, profits will be unusually high and production will tend to increase.

When gross national output is 100, consumers normally may be expected to buy, say, 70 of this total for immediate consumption. Of the remaining 30 units of gross output, capital facilities comprise 15 and governmental services and goods produced for governmental purposes comprise the remaining 15.[2] Since consumption expenditures by the public at large fall considerably short of the value (or cost) of gross national output, it is apparent that huge losses would be suffered if producers had to depend upon consumer demand alone; that is, producers would sell for 70 an output costing 100. The question arises: "How are the other 30 units of cost returned to producing enterprise to prevent such losses?"

First, it may be assumed that government expenditures account for 15 units of the total national output, an amount equal to its tax revenues. Therefore, the funds collected as taxes are used by government to pay wages, salaries, etc., required for the production of its part of total output and for the purchase of commodities (e.g., war goods); government expenditures equal tax collections in this illustration.

There remain the 15 units of output in the form of capital facilities and other investment goods. How are the costs of this output covered or returned to producers? This is done by way of investment of savings in these investment goods. That is, the savers themselves may use their savings to buy capital facilities; other savings may be borrowed by enterprise and used to buy (invest in) capital facilities. Thus investment is the principal way that savings are absorbed, or offset; however, some savings may be

[2] These ratios are approximately the same as those given in Chart 1, p. 101.

borrowed by consumers or by government and spent back into productive channels without going into capital facilities. For the time being it may be assumed that savings are borrowed and used only by enterprise for investment. The important point is that *unless currently saved income and depreciation reserves (gross savings) are invested (or matched by investment) and thus put into productive channels, aggregate demand will fall and with it aggregate supply.* Decisions to invest, therefore, are important for two reasons: (1) they turn gross savings into productive channels, and (2) they maintain and expand the productive facilities of the economy.

It is obvious that without a reasonable expectation of consumption and investment demands, producers of consumer goods and capital facilities will not undertake productive operations. The illustration brings out another point, namely, that if the margin between consumer expenditures and national output (30 in this example) is made larger, the volume of investment and governmental expenditures is of necessity correspondingly increased. Or, to put the matter another way, with a high *propensity to save* there must be a correspondingly high *propensity to invest* in order to maintain a given level of output. This nation saves sizeable amounts each year when output and income are high; some of these savings are invested directly in capital facilities by the savers; the remainder must be borrowed by enterprise and invested if high levels of national output are to be maintained. Investment opportunities are vital to economic welfare.

Producers decide to invest and produce chiefly in expectation of consumer demand. In doing so they create income available to consumers. Part of this income is used for consumer expenditures, part of it is saved by the public at large, and part of it is taxed. The way in which this income is spent, saved, and taxed has a considerable bearing on whether producers realize their profit expectations, and consequently on the production decisions which create the bulk of this income.

PRODUCERS' DECISIONS

The expectation of profit as a motivating influence in economic activity is probably as old as the practice of risking capital, whether it takes such forms as the piratical ventures of the Elizabethan captains or the eminently respectable low-cost housing projects of present-day insurance companies. More than two thousand years ago Aristotle singled out this motive as a threat to the character of Athenian society. Adam Smith, on the other hand, in neatly describing the role of the profit motive, gave it his benediction: "it is not from the benevolence of the butcher, the brewer, or the baker

that we expect our dinner, but from their regard to their own interest." Profit expectation is the *sine qua non* of investment and production. In a predominantly free-enterprise economy, full employment with its high levels of national output and income can be attained only if general social and economic conditions are conducive to favorable profit expectations. As has been noted, the great expansion of the American economy in the past was owing to a combination of circumstances—political, economic, geographic and technical—favorable to the prospects of profit.

Productive activity usually requires an expenditure of money to acquire plant, machinery, equipment, materials, to hire labor, and to cover the other expenses of manufacturing, farming, retailing, and other processes. These expenditures are made usually only if the producer is reasonably assured of success in his venture. For a producer, whether he is a farmer, a merchant, or a large corporation, success depends on (1) "getting back what he puts in" and (2) getting back enough more to make a profit. The expectation of recovering expenditures in a productive venture is clear enough. This is largely what a farmer has in mind when he says his new tractor will "pay for itself in three or four years." Since he knows the life of a tractor is considerably more than three or four years, his statement also expresses the expectation of getting back more than its initial cost. If he did not expect this additional amount, the farmer might use his money otherwise, for example, to buy a government bond paying him interest. To wait three or four years merely to get back the cost of the tractor but no more would, in effect, entail a loss equal to the interest that he would have received from the bond. His expectations, then, are to recover his investment in the tractor plus at least as much as he could have earned in interest. Indeed, he probably would expect even more than the amount of interest as an additional return, because the financial risk, i.e., the possibility of losing money in buying and using a tractor, is considerably greater than the risk in buying a bond.

Suppose the tractor cost $1,000. The risks of farming—a fall in the price of farm products, drought, insect infestation, and other hazards—are such that the farmer might lose the $1,000 entirely through bankruptcy; on the other hand, with very good luck, he might get back as much as $2,000. If he had bought the bond instead he would be assured of recovering his $1,000, plus perhaps $100 in interest. The certainty of getting back $1,100 from the bond purchase must be weighed against the prospects of losing everything at worst, or doubling his money at best, from the tractor investment. He would not risk the loss of his money in buying a tractor if he could expect no more than $1,100; he would risk it if he thought with

luck he might get $2,000.[3] The degree of risk thus looms large in profit expectations.[4]

Prospective Costs and Receipts

The profit expectations of this farmer and all other private producers are formed on the basis of an estimate or guess as to (1) prospective money costs of production for a given period of time, and (2) prospective money receipts from the sale of the things produced during that period of time.

Both costs and receipts may be viewed as *totals*. Total money outlay is compared with total money receipts for a given period of time, and if prospective receipts exceed prospective money costs by some margin, the venture is judged profitable. How large a margin is considered essential will vary from producer to producer, as each regards the risks involved, a "fair" rate of profit for the individual, the customary rate of profit for the particular industry, the size of operations, and other factors. If the prospective margin between total costs and total receipts is less than what the producer regards to be a minimum, all these factors considered, production usually will not be undertaken.

Costs and receipts may also be considered as *averages*. The producer calculates or guesses his average money cost of production for a given period in the future compared with the average price at which he expects to sell his products. If average prospective price exceeds average prospective cost by some margin, the undertaking is regarded as profitable. Again, the size of this margin between price and cost will vary according to the considerations previously stated.[5]

Producers' views on prospective total outlays and receipts, or on average costs and prices, may be based on detailed market analyses that take into account probable future consumer incomes, consumer preferences for competing products, probable results of advertising campaigns, engineering and cost-accounting analyses, and other relevant factors. Large corporations rely heavily on such analyses in making investment and production decisions, although many decisions are based on no more than

[3] He might buy it merely to minimize a loss, but only if he reduced his losses on the farm by more than the cost of the tractor.

[4] It is possible, of course, that some investors are guided by "prestige" considerations. The prestige that goes with having the best tractor or the biggest factory or the latest equipment undoubtedly accounts for some investment expenditures. Many corporations, too, carry out an investment policy year by year with more emphasis on the annual return than getting their money back on each outlay. Nevertheless, when prospects are bleak these investments cease.

[5] In Chap. 14 a detailed analysis of the unit costs of production of the firm is given.

a rule-of-thumb guess derived from present profit experience and a hope that the future will hold no evil surprises. At best, any appraisal of future profit prospects rests in large measure on faith, because future events can be foreseen only imperfectly and often not at all. Risk, in short, is inherent in all investment and production for future sale.

The terms *prospective money costs* and *prospective money receipts*—either as totals or averages—are quite general. Actually, producers can form some idea of these quantities only when they formulate answers to a number of specific questions. Before he undertakes risks a corporation executive, a farmer, the corner grocer, or any other producer will try to answer these questions:

1. *What is likely to happen to the demand for his product, or line of products?* What kinds of things will buyers want in the next few months, few years, or few decades? Are they likely to continue buying the same goods and services as before? Or are new demands developing and existing ones going out?
2. *How much will buyers be willing to buy?* Will they buy as much of his products as before? Or are new demands developing which promise to take large quantities of new products?
3. *What prices will buyers be willing and able to pay for existing and new products?*
4. *How much will it cost to produce these things?* Are technological changes in prospect that will give to him special advantages over other producers, or to them over him? What will labor and raw materials costs be in the future? What effect will governmental policies, especially tax policies, have on costs in the future?
5. *What share of the market will he get?* What kind of competition will he face from other producers?

In analyzing both future expenditures and probable future receipts it is useful to distinguish between the short-term and the long-term calculations and expectations. In general, decisions to produce consumer goods and services—and the materials required for their production—are based on short-term decisions to invest in capital facilities on long-term calculations and expectations. Expenditures for materials and labor to produce consumer goods may be termed *operating* or *working* capital outlays, while expenditures for capital facilities are *long-term investment* outlays. The volume of working capital expenditures, therefore, depends largely on short-term profit expectations; the volume of long-term investments depends largely on long-term expectations.

Short-Term Expectations

The producers of consumer goods, therefore, look to the probable demand of consumers in the near future to guide their working capital outlays and the volume of output. Bakers, for example, expect to sell their bread and other bakery products within a day; clothing manufacturers, their suits, dresses, and other items within a few weeks or months; automobile manufacturers, their cars within two to four months.

Manufacturers of consumer goods gear their production schedules for the period ahead largely on current purchases by consumers (which are taken from inventories on hand) and on forecasts of probable demand (which will be satisfied from current production). If current buying is heavy, inventories become low, and current production may be increased to maintain the inventory, or to increase it. If future demand is expected to exceed current buying, producers in many lines not only adjust production upward but increase their prices as well. On the other hand if consumer purchases fall off both production and price are adjusted downward. In general, manufacturing industries normally emphasize output adjustments in preference to price changes.

Although some producers of consumer goods, such as automobile or radio manufacturers, are able normally to keep themselves closely informed of sales prospects and gauge production on both present sales and expected demand, many producers of consumer goods, and of raw materials for processing into these goods, cannot adapt output to demand readily. The composition and volume of the output of farms, for example, is partly determined by tradition and by climate-soil conditions. Farmers tend to make decisions based on these factors, hoping that the demands for their products in the past will continue into the future. For most of the major crops changes in demand are accompanied by increased or decreased prices with little change in output. Some kinds of farm output, however, are readily adaptable to changes in prospective demand. Truck farmers are able, because of adaptable soil conditions and a relatively small amount of land to cultivate, to vary the acreage planted and types of produce in accordance with expected demand.

The present demand of consumers for goods and services greatly influences the expectation of demand in the months ahead. If demand is favorable at present, there is a tendency to expect it to remain constant or to become more favorable; correspondingly, if demand is unfavorable at present, unfavorable conditions are expected. Present demand, it has been said,

depends largely on the volume of national income payments, and these in turn arise chiefly from current productive activity. Favorable expectations in the short term, therefore, tend actually to create a favorable demand that will lead to realization of these expectations; unfavorable expectations in the short term tend to depress the future market—and thus realize the fears of the producer—by decreasing current production and income payments. These expectations—favorable or unfavorable—are of great importance for they are the bases of producers' decisions to employ workers, produce goods, and pay out incomes. Producers of consumer goods, who produce normally about three fourths of total national output and pay out three fourths of national income payments, largely make decisions on the basis of short-term expectations.

This analysis may be made more clear by the accompanying chart, showing various relationships between future receipts and future costs of doing business. It will be noted that the intersection of the axes is any given moment in time, designated "NOW," and the numbers to the right measure months in the future. The heavy lines showing total receipts (measured in dollars on the vertical axis) measure expected money receipts for each month; the lighter lines measure expected total costs of producing (or selling) amounts of goods which bring in the expected receipts.

In part *A* of Chart 2, the producer (e.g., manufacturer or retailer) looks ahead six months from NOW and expects a growing volume of sales, and possibly higher prices, which promise increasing total receipts. He adjusts his production plans accordingly; if he is a manufacturer he will order more materials, employ more labor, and increase production. This involves an increase in total outlays of working capital (i.e., costs). If a retailer or a wholesaler, he will attempt to build up inventories to satisfy the growing demand of his trade. In any case he will endeavor to maintain as large a spread as possible between total receipts and costs in order to maximize the stream of prospective profits.[6] When expectations in general conform to those shown in *A*, national output, income, and employment increase.

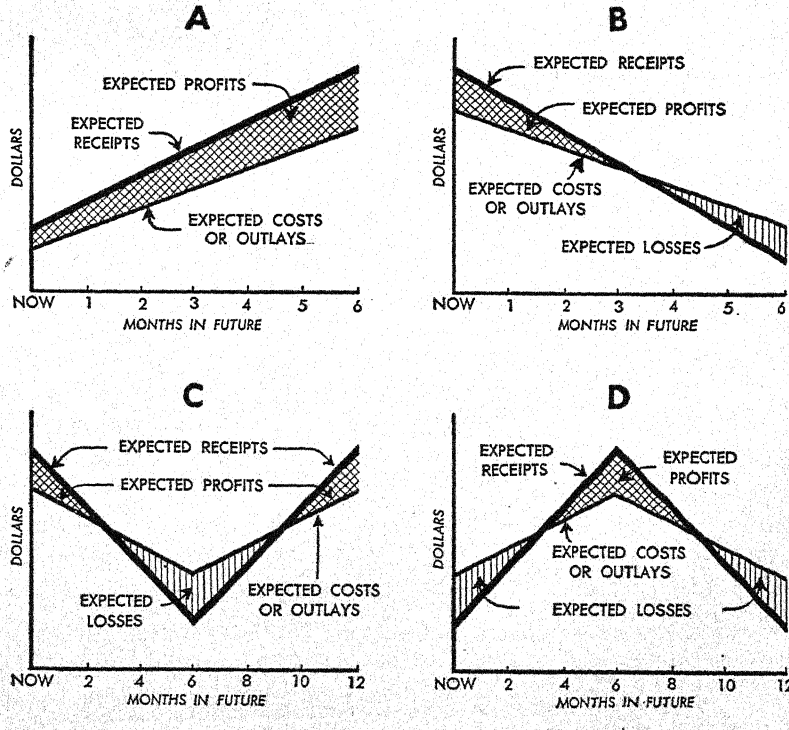
The producer in part *B* faces unfavorable prospects. He expects receipts to fall, either because prices are declining and are expected to decline further, or because at a constant price consumer purchases are falling off, or both. In the face of such bleak prospects, he adjusts his production

[6] The costs relevant here are direct costs or "out-of-pocket" costs. Other costs (indirect or overhead costs) do not affect his calculations, they merely offset part of what has been termed *profit* in the graphs. The fact remains that the producer in any case will attempt to maximize the spread between these curves.

plans (or his purchase plans if a retailer or wholesaler) downward in an effort to keep total costs below total receipts. In the chart this is not accomplished; the producer, however, will endeavor in such a case to minimize his losses, i.e., to narrow the spread where the cost curve exceeds the receipts curve. In some cases, however, a producer may be able to reduce his operations enough to keep costs below revenue and maintain a rate of profit. When expectations conform to the pattern shown in part *B* producers as a whole curtail output, income payments, and employment.

Part *C*, covering a 12-month period, shows expectations during the first 6 months to be unfavorable but those of the following 6 months to be favorable. Producers therefore reduce their operations initially but plan to expand production and sales in the final 6 months. In part *D*, the first half of the year holds favorable expectations and the second half unfavorable. The four situations described in Chart 2 by no means exhaust the possibilities. They are sufficient to illustrate, however, the relationships between expected sales and costs and the streams of profits or losses that,

Chart 2 SHORT-TERM BUSINESS EXPECTATIONS



in a sense, flow through the minds of producers as they contemplate the future and make their decisions accordingly. Every effort will be made by producers to increase the spread between receipts and costs when receipts exceed costs, i.e., they will attempt to maximize the stream of profits. Conversely, when expected receipts fall below costs they attempt to narrow the spread and minimize losses.

Long-Term Expectations and Investment

More complicated than these short-term are those long-term expectations that lead to the investment of funds in capital facilities having a life of several years or a decade or more. Producers require capital facilities to satisfy either directly or indirectly the future demands of consumers and, to a much smaller extent, the future demands of government. Capital facilities are purchased—or produced—by business enterprises only because there is an expectation of a future demand for the goods produced with the aid of these capital facilities. Thus, consumer goods satisfy a direct consumer demand; capital facilities satisfy indirectly the demands of consumers because they are essential to production of consumer goods. The demand for capital facilities, therefore, exists because of producers' expectation of a *future consumer demand.*

In contrast to working capital outlays, expenditures made in acquiring these capital facilities—or in producing them—normally are not returned to the investor without the lapse of a considerable period of time. For example, suppose a bakery contemplates replacing its oven, baking pans, mixers, and its furnace at a total cost of $25,000. These items of equipment can give useful service for 15 to 25 years, and in, say, 12 years the baker may expect them to "pay for themselves." They will pay for themselves in that period if over the entire period the demand for bread and other bakery products and their costs of production are commensurable with the expectations of the baker. Bakery products are staple and nondurable items, of course, and rapid changes in total demand are not likely to occur, but can the baker reasonably foresee demand for 12 or more years in advance? A baker must take into account the specific situation in his city: What changes are likely to occur in the city's population which will affect the demand for bakery products? Are industries moving away, drawing some of the population with them? Or are new industries coming in, promising further growth of the city? What changes are likely to occur in the local supply of bakery products? That is, are his competitors likely to increase their baking capacity? Or decrease it? If the baker is not

reasonably assured of these demand-and-supply factors, an investment of this magnitude will not be made. If the prospects are considered good, the $25,000 will be spent for the capital facilities.

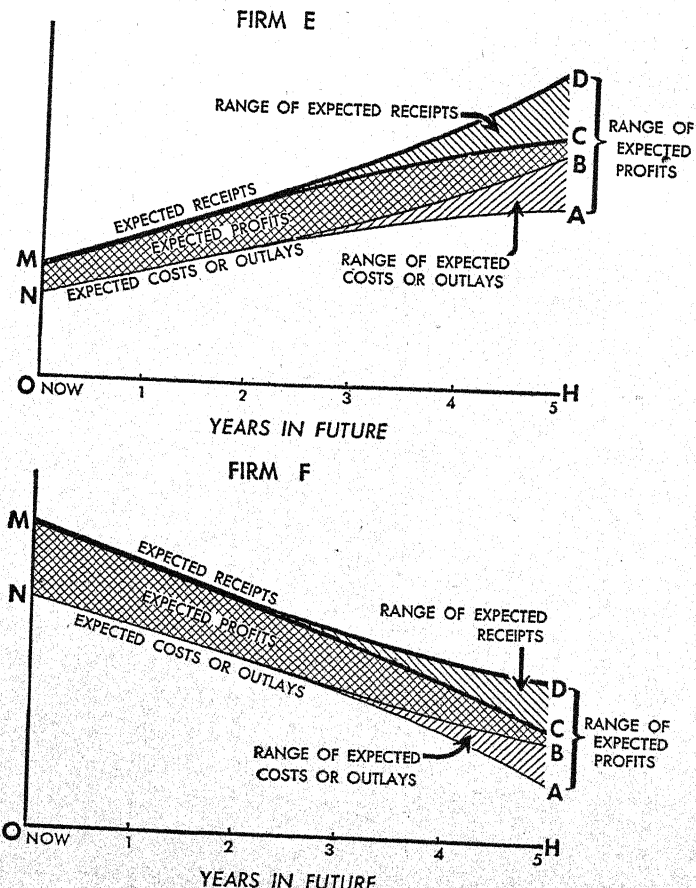
The decision in such a case is made relatively easy because of the traditional stability in the demand for bakery products, and the unlikelihood of any important shift in demand to substitute products in the decade or so ahead. But many industries do not have this stability. The processing of soybeans on a large scale is a relatively new development in the United States, and the future demands for soy products consequently cannot be foreseen with relative assurance by a processing company. A $500,000 investment in soybean-processing equipment, therefore, requires a more sturdy faith and a greater willingness to accept risk than in the case described above. Investment of this sort is aptly described as "sunk" because its recovery must wait many years. On the other hand, working capital outlays are expected to be recovered shortly, certainly within a year. Both types of expenditures involve financial risk, but the risk is far greater for the long-term investment. This is so because, in general, the realization of profit expectations on long-term investment hinges on a sustained or increased demand for a product over a long period of years. Demand in general will be sustained or increased in the future only if expectations continue to be sufficiently favorable to induce enterprise to risk both working and long-term capital and thereby create income. When these expectations are especially favorable output may be such that relatively full employment will ensue; in general, this occurs when substantial amounts of long-term investments are made. When investment is not sufficiently large, output, income, and employment will be below the full employment level.

Chart 3 shows the long-term expectations of two firms. The expected receipts over a five-year period of firm E are shown as rising steadily. However, expectations over so long a period cannot be precise; receipts are therefore shown as a range: *MD* measures the most that can be reasonably expected and *MC* the least. In the fifth year receipts from the sale of output may be as high as *MD* and as low as *MC*, depending upon whether the price of the product and the quantity of output conform to the most optimistic or the least optimistic expectations. The production plans of the firm, entailing expenditures measured from *N* on the chart, are related to the expectations of receipts from the business. The least favorable expectations indicate a profit measured by *BC*, the most favorable by *AD*. In either case the enterprise is prospectively profitable, so that the firm undertakes investment in plant and equipment in order to realize its expec-

tations over the five-year period. It will arrange its production plans to maximize its profits, i.e., to attain the greatest spread between the receipts and cost curves. When producers' expectations generally conform to patterns similar to those shown for firm E investment in capital facilities increases and with it, total output, income, and employment.

Firm F, on the other hand, faces unfavorable prospects. Over the five-year period, total receipts are expected to decline because of prospectively lower prices, or output, or both. In this case, demand for the product in question is declining. In this case also expectations are not precise; receipts may be at *D* or as low as *C*, and the producer will plan output so as to reduce costs in an effort to keep them below receipts. In a case of this type the producer probably would not even replace the capital worn out in the

Chart 3 LONG-TERM BUSINESS EXPECTATIONS



process of production, to say nothing of adding to capital facilities. Indeed, expectations may be worse than those represented by firm F, for some firms may not be able to maintain any profit margin at all. When long-term expectations in general conform to those of firm F, investment declines precipitously, and with it output, income, and employment.

CONSUMPTION, SAVING, AND INVESTMENT

The role of investment and consumption expenditures in the national economy may be described by using again the dollar figures given on page 101, as follows:

Gross national output	$190 billions
Composed of:	
Consumer commodities and services	133
Capital goods	28
Governmental services	19
Commodities produced for government	10

First consider the consumer-goods segment of the economy. In producing this output wages, salaries, and other incomes are paid to those directly or indirectly engaged in its production. These incomes, however, will be less than the value or sales price of this output, for part of this value represents the capital used up in production and indirect business taxes. The tabulation following shows the value of consumer output, the amount of income paid to produce it, and the disposition of this income by those who receive it.

Gross output of consumer commodities and services	$133 billions
Minus depreciation, business savings, and indirect business taxes	20
Equals income payments	113
Consumer expenditures are	92
Individual savings and personal taxes are	21

It is readily apparent that the producers of consumer goods would be sadly disappointed in their expectations if they depended only on the receipt of 92 billion dollars for the sale of an output costing 133 billion dollars. Purchasing power would be short by 41 billion dollars. Where does the additional purchasing power come from?

Here is seen one of the important functions of investment, of capital facilities production. The figures appearing below show the amount of this

output, the income payments made in its production, and the disposition of this income. Income payments to the productive factors are less than the gross output of these industries.

Gross output of capital goods	$28 billions
Minus depreciation, business savings, and indirect business taxes	4
Equals income payments	24
Consumer expenditures are	20
Individual savings and personal taxes are	4

Hence, investment creates a substantial fund of purchasing power which is used largely for the purchase of consumer goods and services. For the two segments of the economy discussed thus far consumer spending is 92 billion dollars plus 18 billion dollars equals 110 billion dollars.

Governmental activity similarly adds to the fund of purchasing power in the hands of the public. The expenditure of 19 billion dollars in the rendering of governmental services goes entirely for wages, salaries and other incomes, and in addition government spends 10 billion dollars to purchase commodities (e.g., war goods, public works). Assume that those who receive income directly from government spend 15 billion dollars for consumption purposes, with taxes and saving accounting for the balance, and the 10 billion dollars of commodity purchases involve the payment of 8 billion dollars of incomes to those producing them. Thus total consumer income and expenditure for the various sectors of the economy are shown in Table 8.

Table 8 SOURCES OF INCOME PAYMENTS

(in billions of dollars)

	Total	*Consumer goods industries*	*Capital goods industries*	*Government services*	*Commodities produced for government*
Income payments	164	113	24	19	8
Consumption	133	92	20	15	6
Individual savings and personal taxes	31	21	4	4	2

Total income payments are 164 billion dollars; consumption expenditures are 133 billion dollars and the difference between income payments and income spent is the amount going into personal savings and personal income taxes. Personal savings are 12 billion dollars and personal taxes 19 billion dollars, or 31 billion dollars as shown above. *When consumer*

spending equals the gross value (cost) of consumer goods and services, short-term expectations in general are realized.[7]

Absorption of Gross Savings by Investment

The value of capital facilities produced is given as 28 billion dollars. It was said above that the production (and purchase) of these facilities depends upon long-term expectations of enterprise. Gross savings are composed as follows:

Individual savings	$12 billions
Undistributed earnings	3
Depreciation reserves	13
Total	**28**

The demand for capital facilities is thus shown to be *equal to the gross savings of the economy.* These gross savings, to repeat, arise from the production and sale of gross national output, and represent a substantial part of the cost of producing this output. As savings, they are not spent for consumption purposes, nor are they used to pay taxes; as savings per se they are withdrawn from the stream of current expenditures.

These gross savings must be used, i.e., spent into productive channels, if the level of economic activity is to be sustained in the period ahead. So far, it has been assumed that they are spent for the purchase of capital facilities, i.e., investment. Thus the flow-back of funds into productive channels is such as to purchase consumer goods output and capital facilities at their cost.[8] Investment, therefore, absorbs the funds available as individual savings and as savings and depreciation reserves of business. This process is sometimes referred to as "offsetting" gross savings, or "putting savings back to work." The significance of investment can perhaps be made clear in this way: Gross savings in part are saved out of the income paid out in the production process by producing units (personal savings are a case in point); the balance of gross savings is retained by producers from the receipts derived from the sale of output. If these gross savings are not spent entirely, i.e., if the demand for capital facilities falls below the assumed 28-billion-dollar level, the production of capital facili-

[7] Gross value equals factor costs (including profit), indirect business taxes, and depreciation. Therefore when consumer spending equals gross value, profits are earned in the production process.

[8] And the flow of taxes into government is assumed to equal the incomes paid out by government, or the value (cost) of governmental services and the purchases of commodities by government.

ties will not continue in the period ahead at the 28-billion-dollar level. The reason for this is clear enough; producers will not produce these facilities when they are unable to sell them (or use them) at a profit. This means that investment falls below the 28-billion-dollar annual rate, and with it, output, income payments, consumption, and savings.

If, on the other hand, investment in the period ahead exceeds the present rate of gross savings, production of capital facilities will increase, and with it total output, income payments, consumption, and savings.

One of the important functions of investment, therefore, is to create a volume of current income which in large part is used for current consumption. When the production of capital facilities falls to a low level during a depression, income payments from this source fall too, and this causes a reduction in consumption and in consumer goods production. On the other hand, when capital goods production is high, the income paid out in this activity stimulates high levels of consumption and output. Traditionally in this country the level of investment has been, and remains, the decisive factor governing the extent to which economic resources are used.

The other, and basically the more important, function of investment is to maintain and build up the productive capacity of the country. As capacity is built up, higher levels of consumer goods production become possible and therefore higher standards of living may be attained; but this capacity will be fully used and higher standards attained only if consumers have the purchasing power to buy consumer goods output at profitable prices. As seen above, investment in capital facilities provides both a large part of this purchasing power and the productive capacity necessary to satisfy the wants of the general public.

In the discussion above, investment was referred to as a means of absorbing or offsetting savings. This need not be interpreted in a static sense. Savings are not accumulated first and then offset or absorbed by investment. Both investment and savings go on simultaneously. Investment, along with other productive work, creates the income of the nation, and savings are set aside out of this income. It is not to the past relationship of savings and investment that one must look; rather the important point is that, for each period ahead, there be a correspondence between investment decisions and the propensity to save.

Governmental Output and Payments

The share of national output produced by government is measured by the wages and salaries paid to public employees engaged in such activities as public education, public health, police and fire protection, postal services, and in the vast administrative, judicial, and legislative tasks performed by federal, state, and local governments in carrying on the many functions of government. It has been assumed above that this totals 19 billion dollars annually, measuring the income payments made by government and the part of national output attributable to government. In addition, government purchases 10 billion dollars of privately produced commodities, so that total government expenditures—for services and commodities—are 29 billion dollars.

It has also been assumed that personal income taxes total 19 billion dollars and that indirect business taxes are 10 billion dollars; thus taxes and government expenditures are assumed to be equal. These taxes are drawn from the current income of the people at large (directly as income taxes and indirectly as costs to business), and consequently are not available to them for either consumption expenditure or for saving. These funds are put back into the income stream when government expenditures are made.

The 29 billion dollars of taxes therefore become income again when spent by government as salaries of public employees, and as wages, etc., paid in the production of commodities sold to the government. It may be said that governmental expenditures offset taxes, or restore to the economy the money withdrawn from the income stream by taxes. Indeed, government may absorb some part of savings as well, by borrowing these savings and spending them. This is done on a large scale during war, and may be done during periods of depression.

Consumption, National Output, and the Volume of Investment

This discussion leads to a point of considerable importance, namely, the proportions of total output accounted for by consumption and nonconsumption purposes. These proportions are of interest for they indicate the magnitude of investment and governmental outlays required to attain given levels of national output.

Thus, when gross output is 190 billion dollars and 133 billion of this is currently consumed, it follows that investment in capital facilities and governmental expenditures together equal 57 billion dollars. Then aggre-

gate demand is equal to aggregate supply and all producer costs (incomes, etc.) flow back to productive channels. Whether these nonconsumption expenditures will continue to be this magnitude depends upon long-term expectations of enterprise and the currently accepted purposes and activities of government. Since the productive role of government is relatively restricted in this country it falls chiefly to private industry to maintain a sufficiently large volume of investment expenditures for high levels of output. This is reflected in the illustration; 28 billion dollars annually must be invested in capital facilities to assure the production of 190 billion dollars of output. This, in fact, will be done when the long-term expectations of producers are sufficiently favorable to induce them to (1) replace worn-out capital facilities amounting to 13 billion dollars and (2) expand capital facilities in existing or new productive activities to the extent of 15 billion dollars.

Suppose, now, with the same gross output, consumption expenditures aggregate 140 billion dollars, instead of 133 billion. Savings will be less and possibly taxes, too. It is evident that less investment is needed now than before (and savings are correspondingly less) to produce 190 billion dollars of output; and governmental expenditures may also be less. On the other hand, if aggregate consumer demand were only 125 billion dollars, taxes and gross savings would be 65 billion dollars; the lower propensity to consume would be accompanied by a correspondingly larger volume of nonconsumption expenditures to produce the given amount of national output.

The proportion of gross output devoted to consumption purposes may vary considerably among nations, or in one nation over a period of time. Great Britain has been referred to as a high-consumption economy because in prewar years some 85 to 90 per cent of its output was in the consumption-goods category. The United States has been referred to as a low-consumption economy, with normally about 70 to 75 per cent of its output in the consumer-goods category. In prewar Russia, Germany, and Japan the percentage of output destined for consumption was perhaps lower than that of this country, for each of these countries devoted a large part of its resources to capital investment and military preparation.[9] The point of importance is that, in order to attain a prosperous condition, a low-consumption economy must invest relatively more, or its government must

[9] These countries, together with Britain and other countries heavily damaged by war, can rebuild and eventually expand rapidly only by holding the consumption ratio at relatively low levels (i.e., increasing the savings ratio), for in this way economic resources in large amounts can then be devoted to capital formation.

spend relatively more, than a high-consumption economy. The high-consumption economy in the short term can maintain relatively full employment with low levels of investment; a low-consumption economy cannot.

The Propensity to Consume

In the section above, consumption and nonconsumption were discussed in relation to gross national output. A closely related matter is the disposition of total income payments among consumer expenditures, individual savings, and personal taxes. As shown in the previous chapter, the propensity to consume tends to decline as income payments increase, and rises as income payments fall. When national income payments were 47 billion dollars in 1933, 99 per cent was spent for consumption purposes; in 1946, income payments were 177 billion dollars and 86 per cent was spent for consumption. Another way of stating the matter is to say that savings, or savings and taxes combined, account for a smaller percentage of low income than of high income.

The percentages spent and saved out of a given volume of income, of course, may change over time. Savings and consumption habits may differ from one period to another, and taxes may take more or less out of a given volume of income at different periods of time. For a period of a decade or several decades such changes may be relatively small; on the other hand widespread social reforms, and particularly the exigencies of war, bring about substantial modifications in savings and consumption habits and in taxation. The general nature of these changes will be discussed in the following chapter.

In concluding this section, the importance of the propensity to consume and its corollary, the percentage of income saved and taxed, should be stressed again. If the propensity to consume is relatively low a nation can achieve and maintain high levels of national output and national income only if private investment and governmental outlays are large; with a higher propensity to consume, these high levels of output and income might be achieved and maintained by a smaller volume of investment and governmental expenditure.

Private Investment and Governmental Expenditures

A few points may now be made on the respective roles of private investment and governmental expenditures in the production of national output.

Is it a matter of indifference how the nonconsumption segment of aggregate demand is divided between governmental and investment expenditure? Are private investments and governmental expenditures equally efficacious in the maintenance of national output? The controversy over large-scale governmental expenditures indicates a real interest, not indifference, in this matter. This is so for several reasons. First, as governmental expenditures increase, taxes in the course of time tend to increase too; some taxes, particularly those that increase production costs and those that rest on profits, might affect profit expectations adversely. Second, large-scale governmental expenditures may tend to put government in spheres of activity traditionally dominated by private producers. This possibility exists because the ordinarily limited field of governmental activity might be insufficient to provide scope for large public expenditure to maintain national output.

However, when private investment in capital facilities cannot be maintained at a rate corresponding to the rate of savings of a full-employment economy, the alternatives seem to be: (1) an increased expenditure by government in fields traditionally private in scope, or if this is not politically feasible or economically desirable, then (2) increased governmental expenditures in directions that do not add greatly to the productive capacity of the country, although they may have much to commend them on other grounds. In this latter category are found such expenditures as payments to veterans, the aged, the unemployed, health benefits, expenditures for the construction of recreational facilities, public monuments, and the like. Failing either of these courses, national output would remain below the full employment level.

Considerations of this sort will not arise to perplex the public and the politician if private investment opportunities are such as to excite favorable profit expectations and therefore induce a large volume of private-investment expenditures. Under these circumstances governmental functions and expenditures tend to be restricted and economic resources are devoted to a very large extent to the development of profitable productive activities.

As long as the laws, customs, and sentiments of the country assign to private producers the chief responsibility for producing national output, the bulk of investment expenditure will be made by private producers guided by their expectations of profit. Whether this amount is sufficient to assure relatively full employment of economic resources depends largely upon (1) the future strength of the underlying economic trends that encouraged favorable profit expectations in the past and (2) the policies of government which bear on future expectations.

CHAPTER 6

Changing Levels of National Output and National Income

In the foregoing discussion it was assumed that the income spent for consumption purposes equals the value (cost) of consumer-goods output, that gross savings are fully absorbed by private investment, and that the expenditures of government equal the tax revenue collected by government. In its actual functioning, the economic system is far more complicated than assumed thus far. Productive activity is largely governed by producers' expectations of profit. The appraisal of future profit prospects is difficult at best; the uncertainty of appraisals is such that the productive activity of the nation is subject to great changes over a period of time. It is the purpose of this chapter to analyze the conditions or factors that bring about these changes. Levels of output and income payments are subject to periodic expansion and contraction commonly known as "business cycles," and to the upheavals caused by war and postwar conditions. The level of gross national output is governed by the aggregate demand or expenditure for this output; the components of this aggregate demand are (1) investment expenditures by private enterprise, (2) consumer expenditures, and (3) governmental expenditures. For present purposes it can be said that increases or decreases in gross national output and national income payments may be initiated in some instances by *investment*, in others by *consumption*, and in still others by *governmental expenditures*. It is perhaps true to say that in the long course of industrial development, both in the United States and Europe, the dynamic changes in national output have been more frequently initiated by expansions and contractions of investment than by changes in consumer or governmental expenditures.

A FALL IN NATIONAL OUTPUT AND NATIONAL INCOME

The following analysis is based (as in the previous illustration) on figures valid for a relatively full use of labor, capital, and other economic resources; and it is developed from a hypothetical period, the Year One. The aggregate demand (or expenditures) for commodities and services in Year One is:

Consumer expenditures	$133 billions
Investment expenditures	28
Governmental expenditures	29
Total expenditures (gross national output)	**190**

The consequences of a decline in investment, consumption, and governmental expenditures, separately and in this order, are described for Year Two and subsequent years.

A Fall in Investment

The volume of investment assumed for Year One reflects the long-term expectations held by private enterprise and measures the demand of enterprise for capital facilities—both for replacement and for expansion—based on these expectations.

With the dawn of Year Two, suppose a change occurs in the conditions which occasioned this rate of investment, making profit prospects appear less certain than during Year One. Since the profit expectations of Year One brought about the 28-billion-dollar volume of investment given above, the new state of expectations will require a smaller volume of investment.[1] That is, business confidence deteriorates during Year Two. Some producers will postpone investments scheduled to be made in the early part of Year Two. Suppose the total amount postponed is such that the *annual rate* of investment early in Year Two drops to 25 billion dollars; i.e., if this reduction persists through the year total investment will be 3 billion dollars below Year One. (Actually, the reduction will be greater than this because a drop in investment tends to cause additional declines in the subsequent period.) The immediate consequences of the initial drop in investment (assuming that it is not offset by increased consumption and government expenditures) early in Year Two will be as follows:

[1] The reduction in expected yield (marginal efficiency of capital) is not offset by a decline in the rate of interest sufficient to maintain the Year One rate of investment.

1. Gross national output will decline immediately.
2. Employment will decrease in the industries producing capital facilities, with a resultant decrease in wages, salaries, and other income payments.

A decline in consumption expenditures will follow from the drop in income payments, because with less income consumers will spend less than before. This means that (1) fewer goods are bought if prices remain unchanged, or (2) if prices fall, as much may be bought as before, or (3) both prices and quantities bought may fall. In all cases, consumer expenditures are less, and approximate an *annual rate* of, say, 131 billion dollars early in Year Two, or 2 billion dollars below the previous year. Consumer expenditures are now less than the value (cost) of current consumer-goods production, which was 133 billion dollars in Year One. When consumer expenditures fall below the value of consumer-goods output, or when the supply of consumer goods exceeds the current purchases of these goods, losses are incurred by producers or their profits are lower than expected.

Short-term expectations of profits being disappointed, producers of consumer goods and services are led to curtail their output and to postpone capital replacement and plans for capital expansion. When the producers of consumer goods see their business declining they tend to expect further declines and, consequently, see no need for capital investment. Thus, gross output declines still farther, and with it employment and national income payments. These additional reductions in output and national income payments will lead to a continued worsening of both short- and long-term expectations in the subsequent months of Year Two, resulting in further reductions in output, employment, income payments, and consumption. The initial decline in investment expenditure in Year Two starts a sequence of reductions in consumer expenditures and long-term investment. This sequence may continue for many months, or for several years, bringing national output, national income, income payments, consumption, gross savings, investment, and personal taxes down to levels below that of Year One.

It is important to understand the interactions of these changes. The initial decline in investment and the consequent fall in wages, salaries, and other income payments affects adversely a wide range of other activities, particularly the consumer-goods industries. The fall in consumption expenditures following the decline in income payments is termed an *induced decline in consumption.* The decline in consumption in turn causes a reduction in the investments made by the consumer-goods industries; this may

be termed an *induced decline in investment.*[2] The decline induced in the output of consumer goods and services is larger than the initial and induced declines in investment when both are measured in terms of dollars. That is, a decline of one dollar in investment might eventually entail a drop of two or three dollars of consumer goods output.

Since an initial decline can cause a general downward spiral of economic activity, it can be assumed for purposes of illustration that the cumulative decline continues through Year Two and Year Three, so that Year Four shows a gross national output (and aggregate demand) of only 140 billion dollars, with employment and levels of income payments, consumption, savings, investment, and personal taxes below that of Year One. This reduction in aggregate demand is assumed to have been initiated by a drop in long-term investment; and the full amount of the reduction in gross output (50 billion dollars) is assumed to have occurred wholly in the nongovernmental sector of the economy, i.e., in the production of capital facilities and consumer goods and services. Government expenditures are assumed to be unchanged. Year Four is compared with Year One in Table 9.

Table 9 COMPOSITION OF AGGREGATE DEMAND

	Year One	*Year Four*	*Amount of change*	*Per cent change*
Consumer expenditures	$133 billions	$105 billions	—$28 billions	—22
Investment expenditures	28	6	— 22	—79
Government expenditures	29	29	0	0
Aggregate demand (gross national output)	**190**	**140**	**— 50**	**—26**

A Fall in Consumption

Consumption can also be the initiating factor in the decline in output, employment, and income payments. In Year One consumer expenditures were 133 billion dollars. Assume that the rate of consumer expenditures early in Year Two falls to 130 billion dollars with no offsetting changes in investment and government expenditures. What would be the immediate consequences?

This change means a shift of income from spending to saving. Since no

[2] Some induced declines in investment may occur directly as a result of the initial drop in investment. Thus the machine-tool and equipment industries which feel the first impact of the initial drop in investment will probably curtail their replacement of facilities and postpone expansion plans.

compensating increase occurs in investment or in government expenditures, aggregate demand will fall. The producers of consumer goods and services will, therefore, suffer losses, or profits will be less than expected. Consequently short-term expectations are not realized and decisions to produce consumer goods will be altered.

Consumer-goods output accordingly will decline soon after the drop in consumer purchases is felt because production schedules are geared to demand, both current and prospective. At the same time, the consumer-goods industries undoubtedly will postpone some capital replacement, and perhaps some planned expansion of capital facilities as well, because the decline in current demand would probably discourage long-term expectations. Thus there occurs an induced decline in investment. Together these reductions in consumer output and investment will bring about a decline in employment and in wages, salaries, and other income payments. This reduction in income payments occasions a further drop in consumer expenditures and this, in turn, a fall in both consumer-goods output and in capital facilities and other investment.

A decline in consumption causes the same downward spiral in an economy as a decline in investment; indeed, it makes little difference in the end result whether the initiating factor is a drop in consumption or in investment. Again, it may be assumed that in Year Four gross national output has fallen to 140 billion dollars; Table 9 may serve for this part of the illustration. As before, it is assumed there is no reduction in governmental expenditures, for government is not ruled by the same motives or guided by the same expectations that regulate the volume of production in the private sector of the economy.

A Fall in Governmental Expenditures

In the hypothetical Year One, government was assumed to have spent 29 billion dollars and to have collected 29 billion dollars in personal and indirect business taxes. Assume that at the outset of Year Two the expenditures of government are cut to an annual rate of 25 billion dollars. Conceivably the governmental employees discharged could be hired immediately by private enterprise (e.g., because of a rise in the rate of investment expenditures), and hence no reduction need occur in national output and national income payments. Assume, however, that no such transfer occurs, i.e., the reduction in governmental expenditure represents a net reduction in aggregate demand.

With a reduction in income payments and gross output early in Year

Two the economic consequences in this case are similar to those already described. Of course, if taxes are reduced by the same amount, consumption expenditures and personal savings may remain unchanged, and likewise the private output and employment provided in producing this output. This is ruled out in the present discussion. The reduction in income payments will soon lead to a fall in consumer expenditure; this in turn will occasion losses, or disappointingly low profits, among the consumer-goods industries. With short-term expectations thus impaired, the output of consumer goods and services will be curtailed, and these industries will postpone some replacement of capital facilities and some planned expansion of facilities. Again, a series of events will ensue following the pattern described for consumption and investment. The reduction in consumer-goods output and in investment entail a further decline in output, income payments, consumption, savings, investment, and taxes; and the spiral works cumulatively downward. In this instance, too, Year Four may show a gross national output of 140 billion dollars; but the governmental contribution to gross output and income payments is now 25 billion dollars, instead of the 29 billion dollars shown above for Year Four, and the nongovernment output is accordingly 115 billion dollars.

The consequences traced in this part of the illustration do not depend upon the assumption made at the outset, i.e., that government expenditures equal taxes in Year One, and that an excess of taxes over expenditures occurs beginning with Year Two. Even if governmental expenditures exceed taxes in both years, a reduction in these excess expenditures may have the consequences traced above, unless private expenditures—of consumers or business enterprise or both—offset this reduction.[3]

A Summary

From the fall of investment, consumption, and government spending, the decline in gross national output, income payments, and employment may be ascribed to the same basic cause: *a reduction in the flow of expenditures in the economy.* Once under way, for whatever reason, the process of recession tends to be cumulative for a period of months, a year,

[3] In 1937 the Federal government deficit (excess of cash expenditure over cash receipts) was sharply curtailed; private expenditure did not increase to offset this decline in the governmental deficit, and aggregate demand declined. On the other hand, the rate of governmental expenditure declined greatly from 1945 to mid-1947; by the latter period the deficit was converted into a surplus of tax receipts over expenditure. Meanwhile, however, a rise in private investment and consumer expenditure more than offset the decline in governmental outlay, and aggregate demand in the latter period exceeded that of 1945.

or several years. Year Four, as shown on page 130, has a gross national output of 140 billion dollars as a consequence of the cumulative declines described above. To repeat, the components of this output are as follows:

Consumer expenditures	$105 billions
Investment expenditures	6
Government expenditures	29

It may be assumed, now, that income payments in this year are 124 billion dollars; the difference of 16 billion dollars between income payments and gross national output is depreciation reserves and indirect business taxes, and each may be assumed to be equal. The disposition of income is as follows:

Consumer expenditures	$105 billions
Individual savings	8
Personal taxes	11

The difference between aggregate demand—140 billion dollars—and consumer expenditure is 35 billion dollars which is composed of 6 billion of investment and 29 billion dollars of government expenditures. Because investment prospects are poor, industry invests less than its depreciation reserves; and because tax revenues have fallen with total income, government to maintain its expenditures borrows the savings which previously (in Year One) had been used by business for its net investment.

In this year of relatively low income, output, and employment the relationships between consumption, saving, personal taxes, and income payments differ from those of Year One. The propensity to consume is higher and the propensity to save lower in Year Four than in Year One. The figures in Table 10 bring out this usual feature of low levels of income.

Table 10 CHANGES IN INCOME DISPOSITION

	Percentage of income payments	
	Year One	*Year Four*
Consumer expenditures	81	85
Individual savings	7	6
Personal taxes	12	9

Moreover, in Year Four the production of capital facilities is depressed; this component of output in Year Four is about 5 per cent of gross national output compared with 15 per cent in Year One. In periods of low-level activity capital-goods industries normally are more depressed than are the consumer-goods industries. A partial explanation of this phenomenon may

be offered. Capital facilities and other investment goods are made up in large part of durable goods; the stock of such goods in existence at any given time can serve for several years without further additions. When expectations become unfavorable, some replacement of facilities is postponed, and new additions may be stopped altogether. In Year Four 6 billion dollars of investment is assumed; this is less than the replacement provided for in Year One and it probably does not replace the facilities used up during Year Four. That is, no net investment occurs in Year Four, indeed *disinvestment* probably occurs; on the other hand, net investment in Year One amounted to 15 billion dollars. Thus replacement is curtailed and new investment stopped altogether, while the economy gets along with its existing capital.

There is less chance for the economy to rest on its oars, so to speak, in the consumer-goods field. Many consumer wants can be satisfied only by current production—food, for example. Demand for many of these goods cannot be deferred. Indeed, it is this characteristic of many consumer goods and demands that helps to explain why the cumulative decline described above does not continue indefinitely. To put it briefly and simply, consumer demand does hold up, and the production of goods (often at reduced prices) to satisfy this demand means employment and income, and this income makes consumer demand effective. If, at the same time, governmental expenditures are maintained—as assumed in Year Four—additional support is given to consumption. Since the ratio of consumption to national output increases during recession, there tends to be a slowing down in the decline in capital-goods production. This is so because investment is destined chiefly to aid in the production of consumer goods and services. As the latter is sustained, long-term expectations take a turn for the better, and at least part of the stage is set for the revival of investment in capital facilities.

A RISE IN NATIONAL OUTPUT AND NATIONAL INCOME

An increase in national output, and in the income payments derived from its production, may be initiated by a rise in the volume of investment, or by an increase in consumer expenditures, or by a growth in governmental expenditures. As in the preceding analysis the close interrelationship of all parts of the economic system must be kept in mind. Just as a reduction in one component of aggregate demand tends to depress economic activity as a whole and bring about relatively low levels of employment, so an increase tends to generate an expansion throughout the economy.

This analysis begins with the low level of economic activity reached in Year Four, i.e., a gross national output of 140 billion dollars. Unemployment in this year is substantial so that idle resources are available to permit subsequent expansion in real output. The changes occasioned by an increase in (1) investment, (2) consumer expenditures, and (3) governmental expenditures during Year Five will now be traced.

An Increase in Investment

Assume that long-term profit expectations are improved so that investment increases from the 6-billion-dollar figure of Year Four to an annual rate of 9 billion dollars early in Year Five.[4] This revival in the "state of business confidence" adds to gross national output immediately an amount equal to the increment of investment. It is assumed here that no offsetting decline occurs in the other components of aggregate demand; i.e., the increase in investment is an injection of additional expenditures into the economic system. Thus output early in Year Five is at an annual rate of 143 billion dollars; and employment in the capital-goods industries rises as unemployed workers are hired to produce the additional machines, equipment, and other items in the capital-investment category. Wages, salaries, and perhaps other income payments increase as a result of this heightened activity in the capital-goods industries and thus the flow of national income payments rises above the 124-billion-dollar rate of Year Four.

The increase in income will tend to bring about an *induced increase in consumption,* for with more income consumers can be expected to spend more. In Year Four consumers spent 105 billion dollars; the rise in investment and income payments brings about an annual rate of more than 107 billion dollars early in Year Five. With consumers now spending more, an improvement occurs in the short-term expectations of retailers, wholesalers, and the manufacturers of consumer goods. Indeed, consumers early in Year Five buy more than producers currently manufacture, for until the rise in consumer expenditures, the output of consumer goods and the volume of consumer expenditures were equal (105 billion dollars) in Year Four.

The increase in aggregate consumer demand and the favorable expec-

[4] Or alternatively an improvement in short-term expectations might occur which would lead business to increase its inventories. The net increase in inventory constitutes a rise in investment. In either case it is assumed that no increase occurs in the rate of interest to offset the prospects of higher profit.

tations generated by it soon leads to an expansion of consumer-goods output, and consequently to an increase in employment in these industries, in transportation, and in retail and wholesale trades. These industries accordingly increase their wage and salary payments in the course of this increased activity; and the larger income impels their workers to increase expenditures for consumer goods and services.

This illustrates a cumulative expansion or an induced increase in consumer income and expenditure following the initial increase of investment early in Year Five. Once consumer expenditures increase, additional increments of income and consumption arise in the course of time because of the stimulus to the expansion of production in the consumer-goods industries. The disposition of the increased income and the effect of this disposition on the expansion occurring in Year Five must now be considered.

Normally income is used chiefly for consumption purposes, while the balance goes into savings and taxes. In Year Four it was assumed that 84 per cent of the income was spent, with taxes and savings accounting for the remaining 16 per cent. This is the average propensity to consume. The 3 billion dollars additional income created by the initial expansion of investment in Year Five will not be spent and saved in these proportions, for as additional increments are added to income the percentage spent declines. That is, the *marginal propensity to consume* is lower as total income increases. Assume that 66⅔ per cent of the additional income is spent and 33⅓ per cent taxed and saved. The amount spent for consumption stimulates an expansion of consumer-goods production and additional income; of this additional income, 66⅔ per cent is also spent which also brings about another increment of output and income. Succeeding and progressively smaller increments of output and income thus result from the spending and respending of the income created by the initial expansion of investment. This is shown in Table 11.

The initial increase in income (3 billion dollars) leads to 2 billion dollars of consumer expenditure; this expenditure becomes income to those who sell the consumer goods, and they in turn spend two thirds of it, and this expenditure becomes income again. Each stage of spending thus provides a smaller increment of output and income, and as the arithmetic is carried out, the increments eventually become negligible in amount. As a result of this process of successive but smaller increments of output, income, and consumption, an initial investment is associated with a rise in output and income that may be several times as large as the initial investment itself. In the table above this expansion of total output and income is three times the initial investment. This stimulative influence is described as the *multi-*

Table 11

EFFECTS OF INCREASED INVESTMENT ON INCOME AND CONSUMER EXPENDITURES

(in millions of dollars)

(Marginal propensity to consume, 66⅔ per cent)

1. Investment		2. Income		3. Consumer expenditures
$3,000 creates	→	$3,000	which leads to →	$2,000
which creates	→	2,000	which leads to →	1,334
which creates	→	1,334	which leads to →	890
which creates	→	890	which leads to →	594
which creates	→	and so on	which leads to	and so on
		9,000		6,000

plying effect of investment on national output and national income payments. The *multiplier* concept, now widely used in economic literature, is a partial description of the stimulative effect of investment on output and income. It is expressed numerically: when $1 of investment is related to a $3 increase in output, the multiplier is said to be 3, as in the table above.

In the dynamic sense, the multiplier describes the stimulative effects of investment on the output of consumer goods and services, although this concept does not completely describe the process of cumulative expansion. For just as an increase in investment may be expected to induce an expansion of consumer-goods output the latter may be expected to induce additional investment. Thus, the stimulative effect works in both directions: (1) investment stimulates additional consumer-goods output, and (2) consumer-goods output stimulates additional investment. The multiplier concept is usually limited to the first sequence. The second sequence is described as the *principle of acceleration.*

The principle of acceleration may be illustrated as follows: Suppose a certain company produces shoes and uses for this purpose 1,000 machines. The annual capacity of these machines is 500,000 pairs of shoes, and 100 machines wear out each year (10 per cent rate of depreciation). To maintain the production of shoes, the company must invest annually in 100 machines. If the demand for shoes increases to 550,000 pairs a year (because total income has increased as described above), the company must increase its machines to 1,100; it must add to the 100 it purchases for replacement another 100 to permit an expansion of output. The capital expenditures of the company therefore double, from 100 to 200 machines, because of the increase in income and consumer expenditure. The prin-

ciple is stated in an extreme form for purposes of illustration; had the company been able to increase production with its existing capacity there would have been no induced net investment. When expansion from low levels of output occurs firms undoubtedly have excess capacity; however, some will not and must expand, others will have to expand facilities after the rise in consumer demand proceeds some distance. In a period of declining output and consumer expenditure (e.g., Years Two and Three) many firms defer replacement of capital and therefore, when consumer expenditure subsequently increases (as in Year Five), they hasten to restore their productive capacity. Investment thus induced by a rise in consumer expenditure in turn is stimulative via the multiplier effect.

These complicated interrelationships may be summarized in the following way:

1. The initial increment of investment early in Year Five induces (the multiplier)
2. a subsequent increase in consumer expenditure, and this, in turn (the principle of acceleration)
3. induces another increment of investment, with its multiplying effect, etc.

It might be tempting to conclude that this process should go on indefinitely—at least until the economy could no longer expand employment and production. But this is not necessarily the case. As shown in Table 11, column 3, each increment of induced consumption following the initial investment is less than the previous one; and, therefore, each increment of investment induced by successive increments of spending is less than the previous one. That is, the interrelationships between the multiplier and the principle of acceleration are such that any given stimulus brings about an expansion which in time exhausts itself, and in fact, reverses itself. This expansion may be termed the *leverage effect* of the initial investment increase, with the leverage compounded of both the multiplier and acceleration principles.

The general expansion of gross national output, employment, and income payments arising from the upsurge of investment may go on for several years as expectations generally improve. This upward spiral might progress to a level of output and employment corresponding to the relatively full use of economic resources. In Year One, this condition was attained with a gross national output of 190 billion dollars. But in all likelihood the expansion beginning in Year Five must go higher to attain relatively full employment because it may be safely assumed (for the

United States) that (1) the labor supply steadily increases and (2) the productivity or output per worker advances over time. The larger and more productive labor supply may require, therefore, a gross national output of perhaps 200 billion dollars if conditions corresponding to relatively full employment are to be attained in, for instance, Year Seven.

What is required to attain a gross national output of 200 billion dollars? Assuming (as throughout the discussion) that there is no change in government expenditure, the decisive factors are therefore consumption and investment. If it is also assumed that there is no change in the consumption ratio during this period, then in Year Seven consumer expenditures and consumer goods and services will amount to 139 billion dollars. In order to attain 200 billion dollars of gross national output, investment must be 32 billion dollars in Year Seven. This follows from the fact that, given the propensity to consume and a stable level of governmental outlays (139 billion dollars of consumer expenditures and 29 billion dollars of governmental services), the balance of 32 billion dollars must be in the form of investment, or capital-goods output.

If people in general were disposed to spend more out of a given income (i.e., an increase occurs in the propensity to consume), or if governmental services and expenditures were higher than assumed here, the volume of investment needed to attain relatively full employment would be somewhat less than the 32 billions. In the following sections these possibilities shall be discussed; for the present they are ruled out, and the expansion beginning in Year Five can attain the 200-billion-dollar level in Year Seven only if long-term profit expectations are favorable enough to cause an investment of 32 billions.

A Rise in Consumption Expenditures

An initial expansion of consumption expenditures in Year Five would bring results similar to investment expansion. In Year Four, consumption aggregated 105 billion dollars, an amount equal to the value of consumer goods and services produced. Suppose that this situation changes early in Year Five, when consumer expenditures increase to an annual rate of 108 billion dollars. This increase is an *injection of additional expenditures* into the economic system in the form of consumer purchases; it is not offset by corresponding declines in investment and government outlays.

Where does the additional money come from? How do consumers in general get it? These additional funds may arise because of a sudden decline in savings. In Year Four people spent 105 billion dollars and saved

8 billion; conceivably this could shift to 108 and 3 in Year Five. Consumers, too, might use up some of their funds on deposit, that previously lay idle in the banks, to finance this additional spending; or many consumers might redeem their government savings bonds, or sell other bonds and stocks. It is also possible that this increase in consumer expenditure is financed by loans; e.g., finance companies and other sources of consumer credit might make funds available for the purchase of automobiles and other durable consumer goods. Why should consumers increase their purchases by loans, redemption of government bonds, or by reducing their savings? One reason may be the necessity to replace worn-out consumer goods—automobiles, clothing, household goods, etc.—after several years of restricted purchases. Or new consumer products may be coming on the market, or improved and attractive models and varieties of existing products. Advertising campaigns and the greater availability of installment credit may encourage and accelerate consumer expenditures.

Whatever the specific causes, the increase in consumer spending in Year Five injects additional money into productive channels. The immediate effect of this increase in consumer buying is an improvement in short-term profit expectations, for purchases now exceed the previous rate of production. This means a "better market" for goods; sales increase and perhaps prices rise somewhat. There can be expected, in consequence, an expansion in the production of consumer goods; at the same time, the consumer-goods industries can be expected to invest additional funds, either for replacement or expansion of capital facilities, or for increased inventories. These sequences have been described above in the discussion of the principle of acceleration.

This upsurge in consumer expenditure is likely to generate confidence among enterprisers and lead to a substantial expansion of long-term investment, well above the 6-billion-dollar volume of Year Four. Indeed, if the rate of expansion of economic activity in Year Five is to continue, investment must respond in this way. The increase in consumer goods output following the initial rise in consumer spending cannot by itself create enough purchasing power to sustain its own increase because part of the income paid to those employed to produce the larger amount of consumer goods is saved and another part goes in payment of taxes. As taxes and as savings these funds are not available to buy consumer goods. Therefore, these funds must be spent back again into the economic system in the form of investment and governmental outlays.

If these investments in fact increase sufficiently, the cumulative expansion will reach a state of relatively full employment. The conditions assumed

earlier—no change in the consumption ratio and no change in the volume of governmental expenditure—were such that relatively full employment in Year Seven required 32 billion dollars of investment. It was shown that with a rise in the consumption ratio, a smaller volume of investment would be required to attain 200 billion dollars of gross national output.

This point can be brought out by the following illustration. If total income payments in Year Seven are 172 billion dollars, and in one case the average propensity to consume is 0.81 and in the other 0.86, the differences in consumer spending are as follows:

$$\$172 \times 0.81 = \$139 = \text{Consumer expenditures}$$
$$\$172 \times 0.86 = \$148 = \text{Consumer expenditures}$$

When these different amounts of consumer expenditures are shown as components of an aggregate demand of 200 billion dollars (= gross national output) the differences in investment expenditure are evident. These are shown in Table 12.

Table 12 **VOLUMES OF INVESTMENT, GIVEN DIFFERENT PROPENSITIES TO CONSUME**

	Assumption A (Average propensity to consume 0.81)	*Assumption B (Average propensity to consume 0.86)*
Consumer expenditures	$139 billions	$148 billions
Investment expenditures	32	23
Government expenditures	29	29
Aggregate demand (gross national output)	**200**	**200**

In the previous chapter attention was directed to the contrast between a "low-consumption" and a "high-consumption" economy; A and B respectively illustrate the contrast, emphasizing the difference in the relations between investment and consumption. Under assumption A, requiring 32 billion dollars of investment, long-term profit expectations must be more favorable, and investment opportunities more plentiful than is the case in assumption B, requiring only 23 billion, for the economy to attain a level of relatively full employment.

Investment is actually not less important in assumption B than in A; merely less investment is necessary to attain the relatively full use of economic resources.

A Rise in Governmental Expenditures

Governmental expenditures may initiate an expansion of output, income, and employment, as demonstrated during the war years. The assumption of a stable volume of governmental outlays (29 billion dollars annually) may now be abandoned to trace briefly the consequences of an increase in this amount.

Assume, instead, that early in Year Five government expenditures began increasing, with the eventual goal being 34 billion dollars annually. This change in governmental outlays may be caused by an expanded public-works program, national defense, an extension of social security, benefits to veterans, farmers—or a combination of several of these. Another assumption must be made: the increase in expenditures is not accompanied immediately by a corresponding increase in tax receipts. That is, all or a large part of the increase in governmental spending is *deficit spending.* This means that *an injection of additional money* into the economic system accompanies the rise in governmental outlays. If taxes increased equally with expenditures, no net addition to the money flow would occur.[5] The effects of the rise in governmental expenditure vary somewhat according to the purposes for which the expenditures are made. Part of the increase may be paid directly as income to various groups, such as veterans, farmers, unemployed, and aged persons. In this case, money income payments increase, and this leads to an *induced increase in consumption* and its effects. Some of the increase in governmental expenditures may be for the building of public works or for armaments, in which instance the stimulation is felt immediately in the industries that produce the materials and equipment for these purposes, having the effects of increases in investment. Indeed, public works may be referred to as "public" investment.

The general economic expansion thus initiated may continue until all economic resources are in relatively full use. This will be the case if the early increase in production, income, and consumption brings about, as described above, a widespread improvement in long-term profit expectations and, therefore, a substantial volume of investment. In the two previous illustrations, assumptions A and B, the magnitude of investment required for relatively full employment in Year Seven was clearly shown. Now, what is the effect of the larger volume of governmental expenditures on investment needs?

[5] This need not be the case necessarily, but the exceptions would entail many complexities at this point in the discussion.

The effect generally depends on how the government obtains the additional 5 billion dollars it spends. These funds may be derived from higher taxes, or they may be borrowed. It is important to know the source of these funds, whether the funds are raised (1) at the expense of consumption or (2) out of the savings of the public in general.

Assume, first, that the government raises this additional money by imposing higher taxes on the people. These taxes may be expected to reduce the volume of consumption of some taxpayers. Additional taxes and governmental expenditures will involve (1) a reduction of consumption on the part of those paying the tax, and (2) an increase in consumption on the part of those receiving the funds from government, i.e., veterans, the unemployed, the aged, and those employed on public works. Items 1 and 2 could be exactly equal; if so the additional expenditures of government would not change aggregate consumer expenditure but merely transfer consumer purchasing power from one group (taxpayers) to another group (those receiving government payments).

Assume secondly that taxes are taken from savings, or more precisely, from funds that otherwise would be saved. Should this be the case, the taxes and expenditures of government would involve (1) a reduction in individual savings of the taxpayers and (2) an increase in consumption on the part of those who receive the governmental payments. Such a tax policy could reduce individual savings by nearly as much as the additional taxes and increase aggregate consumption.

These, then, are the results of drawing from two sources of government income: one does not alter total savings and consumption at all; the other reduces savings substantially and increases consumption by a like amount. If total consumption and savings are not altered by the additional taxes and governmental spending, it follows that the need for investment to maintain a given level of income is also unaffected. If, however, the additional taxes reduce savings and their expenditure by government increases consumption, it follows that the need for investment to maintain income is lessened.

Government, of course, may borrow the additional 5 billion dollars. By this course, the government draws upon the savings of the public in this amount; and by spending these funds restores the savings to the income stream. Most of these funds, as income, will then be spent for consumption. Thus individual savings are not reduced but are used by government for its additional outlays; therefore, the need for investment by private enterprise to maintain aggregate demand is reduced by this amount. The added governmental expenditures, in effect, absorb 5 billion dollars of

savings which otherwise would have to be invested if relatively full employment was to be continued in the period ahead.

Beginning in Year Five governmental expenditures can then, if properly financed, initiate the upswing. As in the two previous illustrations, an additional flow of money is injected into the economic system, and the spending and respending of this money has a leverage effect on output, income, and employment, all predicated on favorable profit expectations throughout the economy. Should a sufficient volume of investment arise, the economy will reach a level of relatively full employment. As shown above, the amount of investment needed to attain this level depends upon (1) the volume of governmental expenditures and (2) the amount of consumption expenditures.

FLUCTUATIONS IN THE ECONOMY

With the aid of the foregoing analysis the changes in gross national output and in aggregate demand during the last two decades may now be reviewed. A caution, however, must be offered. The actual economic scene is far more complex than the foregoing illustrations, in which it was possible to isolate the initiating changes—investment, consumption, and governmental expenditures. In the actual economy this is not always possible, for causes are often obscured in the tangled mass of economic phenomena.

The latter years of the 1920's were marked by high levels of national output and income, low levels of unemployment. The peak in productive activity for the decade was reached in 1929 when gross national output aggregated more than 104 billion dollars, employment averaged 47 million, and unemployment only 2 million. This may be considered a time of relatively full employment, corresponding to the general conditions (but not to the same magnitudes) assumed to prevail in Year One. Aggregate demand for the nation's commodities and services in 1929 is shown in Table 13.

Gross national output declined approximately 13 billion dollars in 1930 and the Great Depression was under way. Gross capital formation fell about 5 billion dollars, consumer expenditures nearly 8 billion dollars, while governmental expenditures increased very slightly. Thus total expenditures for goods and services dropped off, and national output, income payments, and employment declined accordingly. Whether this decline was initiated by a fall in investment or a reduction in consumption is a disputed point, since the statistical evidence by itself is inconclusive.

Table 13 COMPOSITION OF AGGREGATE DEMAND, 1929

	Billions of dollars	*Per cent of total*
Consumer expenditures	78.7	76
Investment expenditures	16.6	16
Government expenditures	8.5	8
Aggregate demand (gross national output)	**103.8**	**100**

Source: U.S. Department of Commerce.

Once initiated, the decline became cumulative, with reductions in investment and in consumption mutually reinforcing each other, as described for the hypothetical years Two to Four.[6] In 1933 the depths of the depression were reached when total output was approximately 56 billion dollars, or 48 billion dollars below the 1929 level. The fall in the value of the gross national output represented both a decline in prices from 1929 to 1933 and a reduction in the physical volume of commodities and services produced. When the decline in prices is taken into account and the value of national output "corrected," the "real output" of the economy declined about one fourth in this four-year period.[7] The figures in Table 14 show the changes

Table 14 CHANGES IN AGGREGATE DEMAND, 1929 AND 1933

	Billions of dollars	*Per cent change from 1929*	*Per cent of aggregate demand, 1933*
Consumer expenditures	46.3	—40	83
Investment expenditures	1.4	—90	3
Government expenditures	8.0	— 6	14
Aggregate demand (gross national output)	**55.7**	**—46**	**100**

Source: U.S. Department of Commerce.

from 1929 to 1933 in aggregate demand or expenditure for the gross national output, the relative importance of each component, and its percentage change from 1929.

[6] The decline in the early 1930's and the deterioration of profit expectations were aggravated by factors not introduced into the illustration, such as the fall in stock-market prices, declining commodity prices, bankruptcy of enterprises, the widespread failures of banks, the contraction of bank deposits, economic and financial disturbances abroad, and many other influences which adversely affected both long- and short-term expectations.

[7] See pp. 102-03 for a discussion of changes in the money value of output and "real" output.

Employment in this year was approximately 38 million, labor supply 51 million,[8] leaving 13 million or one fourth of the labor supply unemployed. Conditions generally in 1933 typify the period 1932-34; in 1932 the decline had almost spent its force and by 1934 a revival was in evidence.

A fact of great importance to be derived from these tabulations is the large percentage decline in capital formation—90 per cent from 1929 to 1933. Capital formation or investment comprised only 3 per cent of gross national output in 1933, in contrast to an average of 16 per cent in the prosperous late twenties. In 1933, investment was not enough to provide for normal capital replacement—the country was "living off its capital"; in contrast, gross capital formation in the earlier period covered replacement and permitted a net addition to the capital of the country. Consumer expenditure and therefore the output of consumer goods and services, on the other hand, declined less drastically than investment and comprised 83 per cent of total demand and output in 1933. The fact that consumer spending declined less than investment served to "cushion" the general decline of output, in the manner described at length in the hypothetical illustration. Governmental expenditures declined slightly because of the general reduction in output and income throughout the economy; i.e., many state and local governments either encountered trouble in raising money or considered retrenchment an appropriate policy during depression.

The decisive factor—but not necessarily the initiating factor—in the general reduction from 1929 to 1933 thus appears to be the heavy drop in investment. The 15-billion-dollar reduction in income payments directly, and the depressing effects of this income reduction, were transmitted to the consumer-goods industries. The output of these industries declined 32 billion dollars, accompanying the 15-billion-dollar decline in investment. This may be said to be the multiplier in reverse; i.e., a reduction, as well as an increase, in investment is associated with a much larger fall or rise in gross national output and income.

The four years from 1934 to 1937 showed a steady growth in gross national output, reaching a level of 90 billion dollars in the latter year. Again, the sequence of change is obscure—all components of output and all sources of expenditure increased. In this period the Federal government embarked upon a spending policy to relieve hardship and stimulate recovery. Investment increased also in 1934 as industry sought to replace worn-out equipment. Together these increases placed more income in the hands of the consuming public. As a consequence consumer expenditures

[8] This is more than two million greater than 1929; the increase measures the growth in labor supply during the four years following 1929.

rose; this increase was encouraged by a growth in consumer credit for the purchase of automobiles and other durable consumer goods. Once this expansion process was under way it became cumulative, following the pattern described by the multiplier and acceleration concepts. Table 15 shows the components of gross national output for this year.

Table 15 **COMPOSITION OF AGGREGATE DEMAND, 1937**

	Billions of dollars	*Per cent change from 1933*
Consumer expenditures	67	+ 45
Investment expenditures	11	+733
Government expenditures	12	+ 50
Aggregate demand (gross national output)	**90**	**+ 63**

Source: U.S. Department of Commerce.

Although the dollar value of gross national output in 1937 was considerably below 1929 the difference is almost wholly due to prices; when corrected for the lower price in 1937 "real" national output was about the same as in 1929. However, from 1929 to 1937 labor supply increased by 4 million; in the same period the productivity of labor increased so that only 45 million persons were required to produce in 1937 what 47 million had produced in 1929. Increased productivity and a larger labor supply resulted in an average unemployment of 8 million in 1937; to have achieved a condition of relatively full employment therefore would have required a considerable increase in real output in 1937 over 1929 (as shown in Year Seven in the previous illustration). Aggregate demand, however, was less than required in 1937 for high levels of output and employment. Investment expenditures were 6 billion dollars less than 1929, and the increase in governmental expenditures did not make up this deficiency. Since income payments were below 1929, consumer expenditures were below the volume of that year. Following 1937 aggregate demand declined, as did national output, income, and employment; expansion was resumed in 1939 as orders for war goods for Europe stimulated demand, and from 1940 war preparations in this country brought an enormous rise in governmental expenditures.

National Output—The War Years and Transition

The war-born expansion of aggregate demand raised gross national output to 213 billion dollars in 1945. Chart 4 shows this expansion, as well

as the changes in output already described. The upper half of the chart gives the figures for each component of aggregate demand; the lower half shows the percentage of the total for each component. In the early stages of war prosperity the principal stimulus came from private investment; this increased 4 billion dollars from 1939 to 1940. From 1941 governmental war expenditures provided the principal stimulus, while private investment was reduced and consumer expenditures restrained by taxes, rationing, and war savings. As discussed above, the 1930's failed to attain relatively full employment because total expenditures for goods and services were inadequate to call forth full production. Governmental expenditures for war, however, were enormous; indeed, the government's demand for war goods and the increased demand of consumers with their war-inflated incomes exceeded the capacity of the country to produce the output demanded. It was chiefly for this reason that rationing, increased taxes, and war-savings campaigns were put into effect as means of restraining the rise of consumer expenditure. That is to say, the propensity to consume was reduced in the war years by governmental policy in an effort to restrain inflation and to divert economic resources to the production of war goods. Chart 5 shows the changes in the use of national income payments and brings out the wartime changes in the propensity to consume.

The end of hostilities in 1945 brought in its train the enormous problems of shifting from a war economy to a peace economy. This transition will be described here only in terms of the changes in the composition of aggregate demand, i.e., the decline in government expenditures and the rise in investment and consumer outlays. Table 16 gives the figures for 1945 and 1946.

Table 16 **COMPOSITION OF AGGREGATE DEMAND, 1945 AND 1946**

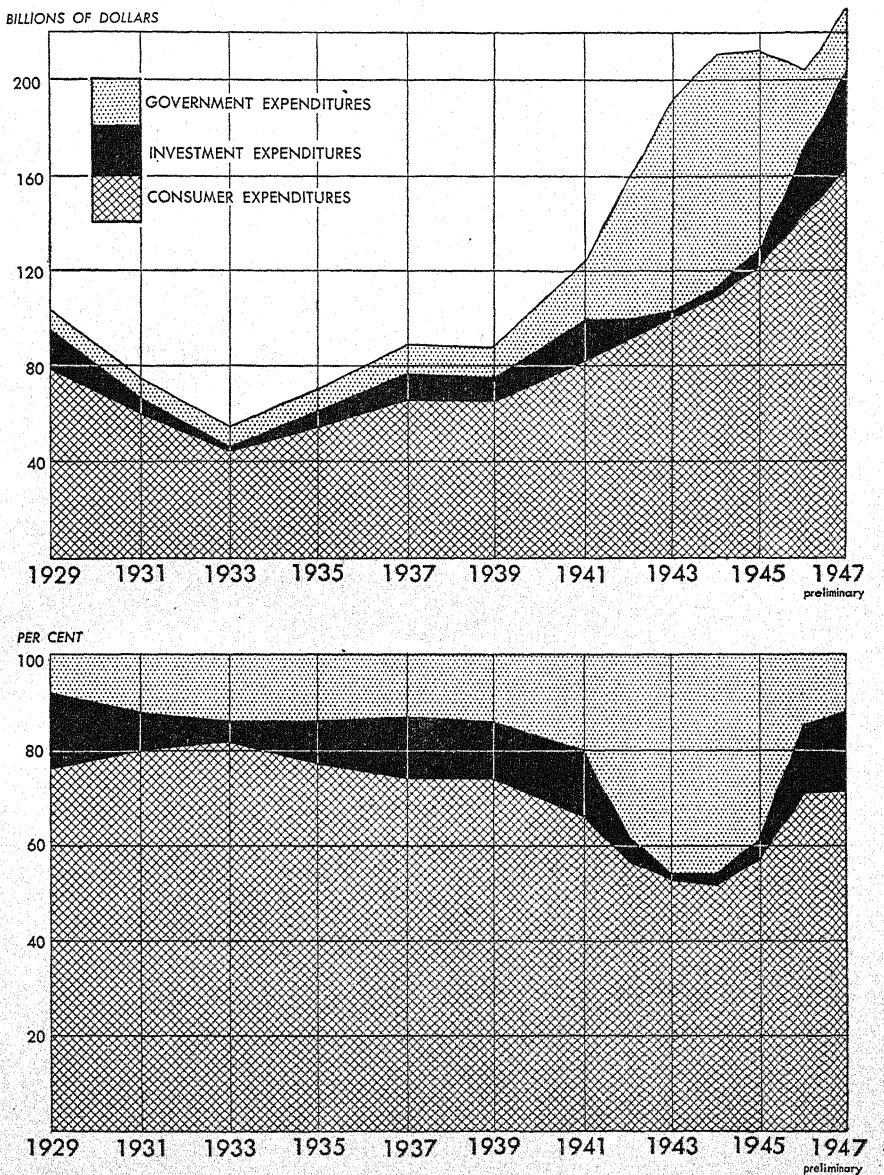
(in billions of dollars)

	1945	*1946*	*Aggregate change*
Consumer expenditures	122	144	+ 22
Investment expenditures	8	30	+ 22
Government expenditures	83	30	— 53
Aggregate demand (gross national output)	**213**	**204**	**— 9**

Source: U.S. Department of Commerce.

The decline in government expenditures in 1946 was accompanied by a rise in consumption and investment. Consumer outlays rose as rationing controls were largely abandoned in 1946 and, with the cessation of price

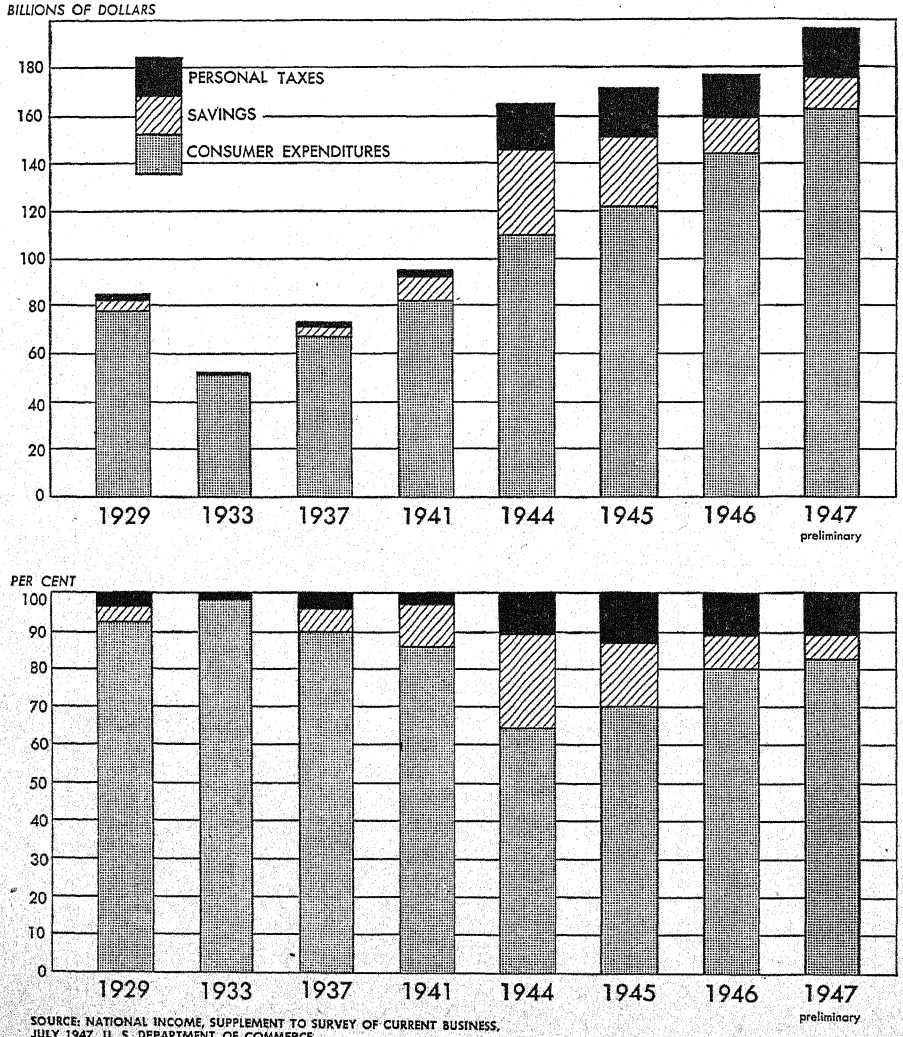
Chart 4 AGGREGATE DEMAND OR GROSS NATIONAL OUTPUT BY COMPONENTS



SOURCE: NATIONAL INCOME, SUPPLEMENT TO SURVEY OF CURRENT BUSINESS, JULY 1947, U. S. DEPARTMENT OF COMMERCE

control, as the prices of consumer goods increased. Saving out of current income also declined in 1946; changes in spending and savings ratios are clearly brought out in Chart 5. The transition from war to peace, therefore, involved great offsetting changes in the aggregate demand for the nation's output of commodities and services; the fall in governmental expenditure was largely offset in 1946 (and more than offset during the first half of 1947) by the efforts of consumers to make up wartime-deferred

Chart 5 PERSONAL INCOME AND ITS DISPOSITION, SELECTED YEARS, 1929-1947



SOURCE: NATIONAL INCOME, SUPPLEMENT TO SURVEY OF CURRENT BUSINESS, JULY 1947, U. S. DEPARTMENT OF COMMERCE

demands and by business to replenish inventories, to replace worn-out capital facilities, and to expand plant and equipment. Shifts in aggregate demand, and in the production of commodities and services, were accompanied by the maintenance of high levels of employment in this period; indeed, so great was aggregate demand in 1946-47 that the peacetime "full-employment" goal was reached and the nation verged on a condition of "overemployment," i.e., a shortage of labor that induced inefficiencies in production, and a shortage of goods in relation to demand which produced inflationary price rises.

The changes in the level of national output and income, reviewed for the period 1929-47, are more complex than those assumed in the hypothetical conditions in the illustration developed in the earlier part of this chapter. The illustration aids, however, in the analysis of the changes that occur in the actual economic system. Aggregate demand—consumer expenditures, investment expenditures, and governmental expenditures—sets economic resources to work; as total demand changes, so does the value of the gross national output. Normally, investment expenditures are the decisive factor in governing the levels of gross national output, income, and employment, so that these levels change as investment changes; and investment varies with long-term expectations. During war, governmental expenditures exercise the dominant role, so that national output, income, and employment change as these expenditures rise or fall. For these reasons, the analysis of the dynamic movements in output, income, and employment placed special attention on the *nonconsumption* expenditures in the economy, for it is on these—investment in peacetime and governmental expenditures during war—that the level of economic activity largely depends.

CHAPTER 7

The Prospects of Economic Progress

The "inquiry into the nature and causes of the wealth of nations" has long been a preoccupation of economic writers. Such inquiry was intensified during the depression of the 1930's in order to examine the causes for

the prolonged "pause" in the country's economic progress. Did this depression signalize the weakening, perhaps the end, of the great expansive forces in the economy? Or were the causes transient, to be followed by a resumption of great volumes of investment and output? Recent times have seen many divergent interpretations of economic development and the rise of several comprehensive theories on the prospects for the future. These several interpretations relate not only to the prospects of economic progress but also to future political and social developments.

First, the areas of agreement among theorists should be outlined before their differences are examined.

1. The material well-being of the country—the standards of living of its people and its position and security in the world—depends upon continued expansion of national output, income, and employment, and these in turn require increasing volumes of investment.
2. Labor supply has increased throughout the history of this country and will continue to grow for a long time to come. Therefore, employment and output must expand for a long time to come if considerable unemployment is to be avoided.
3. The advance and growth of technology have increased the efficiency per man-hour of employment as far back as the records go, and will continue to do so in the future. Therefore, if relatively full employment is to prevail in the future, national output must increase more than the labor supply.
4. Investment expenditures were inadequate during the prewar decade, and this inadequacy accounted for the depressed levels of output, income, and employment.
5. Relatively full employment in the past required a considerable volume of *net* investment to absorb the net savings of the economy. This is expected to be true of the future, too, if the country is to enjoy relatively full employment.

The decisive importance of investment is evident in these points. The issues summarized devolve largely on the future prospects of adequate investment expenditure, and on the conditions that govern these investments. For present purposes the divergent views on investment prospects may be classified as follows: (1) the *mature economy* thesis, (2) the *expansive economy* thesis, and (3) the *capitalist disintegration* thesis.

THE MATURE ECONOMY THESIS

This thesis asserts, as the term suggests,[1] that the American economy has reached a mature middle age. Its vigorous, growing youth is past. It cannot sustain the rate of growth which marked the first three decades of this century, with ever-higher levels of investment, output, income, and employment. The decade of unemployment and depressed investment which followed is viewed as an early but major manifestation of the economy's maturity, and one which will recur in the future. It is held, then, that the great forces which drove this economy to constantly increasing levels of output are exhausted, or largely so. Therefore the system, as traditionally organized and guided, cannot be expected to provide relatively full employment for long intervals. The sporadic energies of a middle-aged economy may lead occasionally to high levels of output, but to sustain high levels of output and employment is beyond the powers of the economy. The conclusion drawn from this point of view is that government regulation and spending are necessary to maintain consistently the welfare of the economy.[2]

The Expansive Forces

A summary of this thesis must begin with a description of the major expansive forces in the economy in its earlier period of development. The basic trends were: (1) an enormous growth in population, (2) geographical expansion and the discovery and exploitation of vast resources, (3) technological changes and the mechanization of industry. These three forces formed the basis of American economic expansion because *they created enormous investment opportunities;* because of them the savings of the economy, and some foreign savings as well, were absorbed by ever-increasing investments. This is not to say that economic progress was steady year in and year out, but despite periodic reversals the expansion was strong and insistent. The central idea in the mature economy thesis is that these forces have been weakened or dissipated. The twentyfold increase in population between 1800 and 1920, it is claimed, called for a great expansion in

[1] This is sometimes termed the *economic stagnation theory.*

[2] The acknowledged leaders in the development of this thesis are J. M. Keynes, whose work *The General Theory of Employment, Interest, and Money* (Harcourt, Brace, 1936) represents the fullest theoretical statement of the matter, and Alvin Hansen, the leading American proponent of this position (see particularly his *Fiscal Policy and Business Cycles* (Norton, 1941).

home construction, in capital facilities to produce food, clothing, and other goods, in schools, hospitals, water systems, and other public works. An expansion of investment, output, and employment was required by the sheer increase in numbers—by the need to increase existing facilities to accommodate the rapidly growing population.

With this growth in population went the westward push of the frontier—space for the growing population. This geographical expansion made room for new communities, for new farms and industries, and required extensive transportation systems. All this called for increased investment. The new areas were rich in natural resources, as noted earlier. The development and use of these resources not only demanded large-scale investment, but also provided cheap raw materials, which in turn encouraged industrial expansion in the settled parts of the country.

At the same time, new inventions created new industries and revolutionized many of the old ones. Rapid changes in technology gave the nation a succession of great industrial advances, led by the development of new industries, such as the railroads, steel, electrical utilities, and the automobile industry. Mechanization spread rapidly, extending from the new industries to old industries. The dynasty of coal and iron, electrical power, and oil emerged from the agricultural-handicraft economy of the early nineteenth century. This transformation required capital for replacement and expansion on a scale that readily absorbed the savings accumulated out of the rising level of income.

Symptoms of Maturity

According to the mature economy thesis, these expansive influences have in the course of time lost their force, and investment opportunities have therefore contracted and the need for savings has been reduced. But the savings habits of the country have not corresponded to this alteration. The public now tends to save as before, but relatively less savings are needed because relatively less investment is possible. This argument of the maturity thesis should be traced in some detail.

The rapid growth of population has come to an end. In the decade 1930-39 population increased 8.9 millions, compared with 17 millions during the years 1920-29. And the population experts predict further slowing up in the rate of increase in each succeeding decade until, near the end of the century, the maximum size will be reached.[3]

[3] See pp. 77-79 for an examination of the changes in the *composition* of a declining or static population and their effects on the economy. The increase in birth rates during

Investment and savings are thus affected in several ways. First, the need for facilities to feed, clothe, house, and educate the population, i.e., investment, tends to slacken as population growth becomes static or declines. For example, the 17 million addition to the population during the 1920's required a greater amount of new housing and schools, new facilities to produce food, clothing, and other goods, than did the 8.9 million increase during the 1930's. This tendency can of course be projected into the future as population increases become eventually smaller. Second, the small-sized family, which is obviously characteristic of a "mature" population, tends to save a relatively larger part of its income than a large-sized family—its propensity to save is greater. A family of three can save more on a $5,000 income than can a family of five. Third, the expenditures made by a small family tend to include more services (in contrast to goods) than do those of a larger family; and service industries require smaller capital expenditures than goods-producing industries.

Thus trends toward decrease in population and reduction in family size diminish the requirement for, and the incentive to make, great volumes of capital investment; the same trends, coupled with attendant rises in per capita income, tend to increase savings. In short, the economy must look toward a future in which the stimulative, expansive influence of a rapidly growing population no longer contributes an important share toward the absorption of savings through investment.

The mature economy thesis also views the end of the frontier and geographic expansion as a depressive factor. The capital required to make new areas habitable and productive is less as the country becomes settled. There remain, of course, large capital needs for the replacement of transport systems and productive facilities in the areas long settled. But the replacement of existing facilities, or even the expansion of these facilities, is not deemed so stimulative as the initial investment made in the period of geographic expansion. For instance, building the transcontinental railroads during the latter nineteenth century gave a greater stimulus to output and employment than does current replacement of rails and rolling stock. The great westward movement came to an end with the close of the nineteenth century, and this economy (say the proponents of the mature economy thesis) never again can recapture the vigor and enterprise engendered by geographic expansion. The lessened need for capital and the diminished investment opportunities arising from "geographic maturity" reinforce those same influences arising from population stability.

the present decade, discussed in Chapter 3, makes necessary a modification of these population forecasts.

The third major element in the mature economy thesis relates to *industrial technology* and *new industries.* Part of the past expansion of this country is attributable, in this view, to the appearance periodically of great new industries. The automobile industry is an example. This industry required large amounts of capital, and caused the expansion of related industries, such as petroleum, rubber tires, and service trades. By the mid-twenties, however, the automobile industry and its related industries had "grown up." Part of the explanation of the depression of the 1930's is the failure of technology to produce a new industry of the magnitude and stimulative effect of the automobile, electrical, or railroad industries.

The mature economy thesis, however, goes beyond this point and suggests that new industries of the present and future can no longer exert the stimulative influence that new industries did in the past. The reasoning is that decades ago a new industry could not be supplied, from existing productive facilities, with the materials and equipment it needed for its growth. Industrial capacity was not sufficiently large or diversified to supply steel, copper, machine tools, equipment, etc. The steel industry and coal mining had to be transformed to supply the railroads; the machine tool, glass, and petroleum industries had to grow and change greatly to supply the automobile industry with the things needed for the production of cars. That is, the basic industries supplying materials, equipment, power, fuel, and lubricants had to be vastly expanded to meet the needs of the "new" industry. By the 1920's these basic industries were well developed, and others, such as aluminum and magnesium, were greatly expanded during the recent war. In the future, therefore, a "new" industry can be well supplied—perhaps fully supplied—with necessary materials and equipment without calling forth an expansion of the existing basic industries. As a consequence, it is believed, the impact on the economy of a new industry will be less powerful in the future than previously because it will not induce comparable expansion in the capital facilities of the collateral or supplying industries. This is important because, in the past, the collateral industries required more capital for their growth than did the single new industry which initiated their growth.

The mature economy view maintains, further, that the character of technological innovation has changed. In the earlier phases of industrialism the innovations were of the "capital-consuming" type, i.e., capital investment in this phase increased more than total output. From the late 1920's, however, the technical innovations were "capital-saving" in character. That is, the trend is toward reducing the capital costs per unit of output. Old machines are replaced by new ones which are cheaper to install and to

operate and which produce a greater output. Also new machines often last longer than the old ones. A railroad company may replace an old locomotive with a new one costing no more than the replaced one did originally, but the new locomotive will pull twice the number of freight cars and use less fuel in doing so. As a consequence of this type of technological change, capital requirements grow less than output.

Population stability, the end of geographic expansion, and technological changes, therefore, limit the need for, and curb the expansion of, capital investment. It is admitted, of course, that the capital replacement needs of the economy are large, but these needs are provided for by the retention of funds in depreciation reserves. However, future expansion of the magnitude required needs more than mere capital replacement—there must be substantial *net* investment as well, to correspond with the net savings of the economy at high levels of income. This net investment of sufficient size is inhibited by the factors described above. The economy faces a deficiency of normal investment opportunities in the future; therefore, it cannot for long periods enjoy relatively full employment.

The concept of a deficiency of investment opportunities and investment expenditures may be expressed more precisely. Suppose it takes 28 billion dollars of private gross investment each year to attain relatively full employment. Of this, 13 billion is replacement and 15 billion net capital expansion. But long-term profit expectations may not warrant this amount of investment; at most, gross investment may amount to only 23 billion dollars. Thus investment is 5 billion dollars short of providing relatively full employment. Or the matter may be put this way: It would take a gross output of, for example, 190 billion dollars to attain relatively full employment. From the income people would get under these conditions, individual savings would total 15 billion dollars, and depreciation reserves of enterprise would amount to 13 billion dollars. That is, gross savings equal 28 billion dollars.[4]

It has been seen that these gross savings must be absorbed by investment if output is to be increased in the period ahead. However, if long-term expectations allow for this period only 23 billion dollars of investment, then investment will be deficient in offsetting the savings that would arise with relatively full employment. This is "oversavings," as the proponents of the mature economy thesis use the concept. It means that *the propensity to save is too high in relation to investment opportunities.*

[4] These figures are taken from the hypothetical model on p. 130–Year One. Consumer expenditures are 133 billion dollars and taxes 29 billion dollars.

The Remedies

The remedies proposed from this analysis of the mature economy can be described here only briefly. They fall in two general categories: (1) proposals for government spending; (2) proposals for increasing consumption and for decreasing saving.

It is proposed, first, that government could spend (or "invest," as the maturity economists prefer) a sufficient amount to make up the deficiency. In the illustration above, an additional 5 billion dollars of governmental outlays would "fill the gap." Thus private investment would be 23 billion dollars, public investment, 5 billion dollars; the total, 28 billion, would assure relatively full employment.

It is proposed, also, that government spending may be coupled with a second measure: a decrease in the propensity to save by the taxing of high incomes. The funds thus received from taxes would not be saved; instead, they would be paid as wages in a public-works program to workers who would spend them largely for consumption. Other means designed to increase the propensity to consume include reduction of taxes on low-income groups, encouragement of high-wage policies, and social security measures to assure minimum consumption standards for the aged and unemployed. All of these proposals are intended to increase consumption and decrease savings. Thus, if such proposals shifted 5 billion dollars from savings to consumption, 190 billion dollars of gross output (under the previous illustration) could be attained with an investment of only 23 billion dollars.[5]

The specific character of these remedies is less important for present purposes than the basic analysis from which they issue. The mature economy analysis purports to demonstrate that fundamental long-term changes in the economic system will limit its capacity to expand in the future.

THE EXPANSIVE ECONOMY THESIS

According to the expansive economy thesis,[6] the expansive influences in the American economy are so strong that, if permitted to operate without hindrance, they will bring about high levels of production and employ-

[5] Presumably the shift need not be so great, for the increase in consumer expenditures would elicit some expansion of investment over the 23-billion-dollar level by means of the acceleration influence.

[6] See George Terborgh, *The Bogey of Economic Maturity* (Machinery and Allied Products Institute, 1945) for a statement of this thesis and a criticism of the maturity view; also Martin V. Jones, *Secular Trends and Idle Resources* (Univ. of Chicago Press, 1944).

ment. Contrary to the mature economy thesis, there are no inherent deficiencies in investment opportunities. It is held that the need for high levels of investment is more urgent and the underlying economic conditions are more favorable to large-scale investment than in earlier times. What, then, occasions stagnation of investment? The answer is: Obstacles are placed in the way of enterprise, or there are deterrents to investment. These deterrents are found in many practices and policies of government, labor, and business. Since enterprise was relatively free of such deterrents in the period of its great expansion, it is believed that normal expansion can continue in the future if enterprise is again allowed to operate without the hindrance of excessive governmental controls and taxation, of burdensome labor demands, and of restrictive monopolistic practices.

This view is a restatement of the traditional faith in free private enterprise as the most effective means of attaining maximum national output and assuring the material welfare of the economy. The mature economy thesis questions the effectiveness of private enterprise in these respects, on the grounds summarized in the foregoing section. In its turn, the expansive economy thesis attempts to demonstrate the fallacy of the mature economy idea.

Criticism of the Mature Economy Thesis

Proponents of the expansive economy position attack the argument that the declining rate of population growth brings with it a narrowing of investment opportunities and a tendency toward oversaving. In the first place, they question the evidence on which the argument rests. The decline in the rate of population growth is not a recent phenomenon. In this country the rate of growth began to diminish after the middle of the nineteenth century, yet this country enjoyed unparalleled expansion of investment in this period. Other industrial countries, too, combined an increase in investment with a declining rate of population growth. On the other hand, some countries with a rapidly growing population are poverty-ridden. If the slowing up of population growth were a depressive factor in this country, why was it not evident before 1930?

Furthermore, the expansive economy proponents argue that a population which tends toward stability will be inclined to save relatively less than an expanding population. In a stable population the aged group which increases numerically and relatively, probably tends to spend previously accumulated savings; in the technical jargon, the aged group *dis-saves*. To illustrate: a retired person may live on an old-age annuity, or on his sav-

ings account, or on war bonds or other securities, all of which represent accumulated savings. Several decades ago less than 5 per cent of the population was sixty-five years old and over, and few had annuities; several decades hence about 15 per cent of the population will be in this age group, almost all of whom will receive annuities under the social security program. Under these circumstances the mature economy thesis is judged inaccurate, since it is held that no excess of savings, in relation to investment opportunities, arises out of population changes.

The expansive economy thesis depends on steadily mounting standards of living in the future, instead of a rapidly increasing population, to provide the basis for large-scale investment. The slowly growing population will need fewer additional houses as time goes on, but it will need better houses to replace the present ones. And to feed and clothe the population more adequately in the future will require increasing volumes of investment and national output. Living standards can be improved, and each generation is expected to set its goal above the previous one.

It may be noted that the desire for an improvement in living standards does not automatically translate itself into increased investment and output; but neither do the needs of a growing population. Investment is inspired not by desires and needs per se but by inducements. These inducements are prospective profits, which impel investors to anticipate the seemingly obvious and certain needs of a growing population. Is it as likely that investors will anticipate needs as readily for a relatively stable population which, however, aims toward progressively higher standards of living? The expansive economy proponents, in opposition to the maturity proponents, believe they will.

The depressing effects of geographic stability are also discounted in the expansive economy thesis. The stimulative effects of geographic expansion upon the American economy are not minimized, but it is pointed out that while the frontier had disappeared by 1900, investment and output continued to expand in the succeeding three decades. If the passing of the frontier created a dearth of investment opportunity, there was no evidence of it. The ratio of private investment to national output was about the same from 1900 to 1930 as from 1880 to 1900.

Contrary to the mature economy thesis, it is argued that after an area is settled increased amounts of investment are required for replacement and for the expansion that goes with the "filling out" process. In fact, the early stages of settlement entail relatively small amounts of capital investment: housing is rough and simple, railroads are single track and of short dis-

tances, industrial plants are small, and municipal facilities are limited. As the settlement progresses the original installations are frequently abandoned or torn down and replaced by structures requiring many times the initial capital expenditure.

The dearth of new industries and the changing character of technological innovations are also dismissed as factors contributing to economic stagnation. The proponents of the expansive economy thesis believe too much importance has been attached to specific new industries by the mature economy theorists. Admittedly the big new industries of the past—the railroad, metallurgical, electrical, and automotive industries—accounted for large investments; but in each period when one of these was supposedly the dominant "new" industry, the bulk of investment came from other sources. Indeed, in each period the new industry was dwarfed by one of the very "old" industries, namely, building construction.

There are other flaws charged to the new-industry argument. The railroads invested more money during the "electrical industry period" than in the era assigned to them by the mature economy theorists; and the electricals invested more during the so-called automotive era than in the period of their alleged supremacy. Moreover, the investment in railroads during the electrical-industry period was greater than in the electrical industries themselves; and the latter in turn invested more funds during the 1920's (the automotive era) than did the automobile industry. Indeed all the current and past "new" industries accounted for less than 20 per cent of total investment made in this decade.

Thus the idea that a new industry of overshadowing importance paced the expansion of investment in certain past periods is given little weight on the basis of statistical evidence. And therefore the economy in the future need not depend upon a single large industry to raise the level of investment to new heights. Instead "Capital formation is a tree with a thousand roots. The great bulk of its nourishment comes always from long-established industries, not from new ones." [7]

Technological limitations and trends do not stand in the way of large-scale investment. At any given time there are many technical innovations and new products awaiting introduction. The rate at which these are perfected technically is usually greater than the rate at which they are introduced into the productive processes. This is particularly true during war because technology is speeded up but commercial application is virtually halted. In the expansive economy thesis, the future holds promise of an

[7] George Terborgh, *Public and Private Investment after the War* (Machinery and Allied Products Institute, 1943).

imposing array of new materials, new techniques, and new products which could create many new industries and revolutionize many old ones; and these changes could absorb great volumes of savings. However, the commercial application or development of these materials, techniques, and products is a nontechnological matter. An economic incentive is necessary to make such development practicable. And the incentive to invest funds for the application or use of these materials, techniques, and products depends in large measure on the policies and practices of business, labor, and government. If these are favorable to investment, full levels of national output, income, and employment will be assured; if not, i.e., if obstacles or deterrents are set up by these policies and practices, a stagnation of investment may prevail.

The Spirit of Enterprise

In substance, the expansive economy thesis asserts that the underlying economic conditions are favorable to expansion, but realization of expansion depends upon a vigorous spirit of enterprise. In the past, enterprise had virtually a free rein in the exploitation of economic opportunities, and the willingness to invest in the development of these opportunities brought economic progress. Investment involved large risks, but enterprisers accepted these, for success was crowned with profit and with the prestige accorded to conspicuous achievement. These rewards inspired the spirit of enterprise. Moreover, opportunities to invest were relatively unrestricted; the fields of enterprise were "open" to those possessing the skill, courage, and money to venture forth.

Three factors underlie large-scale investment: (1) prospective rewards which are commensurate with (2) the normal risks of investment, and (3) a free scope for enterprise. If rewards are cut down, or risks increased, or the scope of activity restricted, investment (and therefore output, income, and employment) will fall below maximum levels. The spirit of enterprise, it is declared, may be stifled because incentives or rewards are inadequate and risks excessive. The failure to achieve full recovery from the prewar depression is explained in these terms. For the future a continuation of economic progress will be assured if incentives to invest are preserved and the risks of investment are minimized, i.e., if the obstacles or deterrents to investment which characterized the prewar decade are removed.

Investment Deterrents

The principal deterrents may be summarized in terms of their incidence on (1) rewards, (2) risks, and (3) scope of investment.

Taxation has the most decisive effects on the rewards or profits of enterprise. Decisions to invest, as noted earlier, involve an appraisal of prospective returns and prospective risks. Prospective returns of necessity are reduced by income taxes, both corporate and personal; the higher the tax, the lower the prospective returns. Therefore, the levy of income taxes may render some prospective investments unattractive. When the tax is low the amount of investment thus deterred will be small; as taxes increase the amount deterred grows. If the tax is 100 per cent of the return, no private investment would be made at all. In an earlier period this country imposed no taxes on corporate and personal incomes. The rewards of risk and enterprise remained with those who got them. The inception of the income tax in 1916 changed this; rewards are now shared by government and enterprise. During the 1930's tax rates increased considerably—the governmental share in profits became appreciable. According to the expansive economy thesis,[8] high taxes contributed to the prolongation of depression, for incentives to invest were weakened.

The reduction of earnings by taxation has a discouraging effect on the growth of small firms and to a lesser extent on large firms. Early in the century the Ford Motor Company could grow to enormous size by plowing back its earnings; taxes now make such self-financed expansion virtually impossible.

The reduction of rewards implicit in income taxation is not accompanied by a corresponding reduction in risk. Government shares in the returns, not in the risks. Enterprise bears the full risk of investment, but receives only part of the return. Thus income taxation is viewed as a deterrent to investment; and the deterring effects grow as tax rates rise.

Other factors, too, obscure and reduce prospective returns on investment. The increased economic and political strength of organized labor leads to recurring—and often successful—demands for wage increases. Governmental support of such demands tends to increase their frequency and effectiveness, and wage issues may therefore be settled not on economic but on political grounds. The prospect of politically encouraged wage demands is deemed disturbing to profit expectations and thus contributes to a weakening of investment incentives.

[8] Many who support the mature economy view also accept this point.

The increase in risks stems partly from the tax and wage problems discussed here. Not only are taxes high, but they are variable from year to year; frequent changes in tax rates and tax provisions increase uncertainty and risk. Furthermore, there are changes in union wage and other demands; governmental intervention in business affairs, governmental reforms and social experiments, government ownership—all increase the businessman's risk. Intervention is regarded as "changing the rules of the game," as a cause of uncertainty in prospective profits. When governmental policy is inspired by an "unfriendly attitude" toward enterprise, uncertainty and risk are magnified. The increase in risk—real or fancied—and the reduction in returns together exert a depressing effect on new investment.

Business itself often restricts investment opportunities. The monopolistic and quasi-monopolistic practices of business are intended to reduce competition. In many fields the "Big Four" or "Big Five" corporations control three fourths or more of the total business; the small "independents" often live by sufferance, with their existence dependent upon supine acceptance of the minor roles permitted them by the big corporations. Small enterprises cannot attempt to expand and continue to live, nor can newcomers easily enter the field. Furthermore, the big companies are often content with their own lots, for if one should expand the others must follow suit. Thus the incentive to expand (new investment) is weakened. Keeping the "independents" in line and prohibiting newcomers may be enforced by control of patents, of raw materials, or by the threat of a price war and by other measures. And self-discipline among the big companies, i.e., each keeping to his accustomed share of the market, is often the price for maintaining respectability among one's associates. The effect of all this is to impose inflexibility in many fields of productive activity; expansion is limited, new productive techniques may be discouraged, and the enterprising spirit is stifled in the interest of "position" and security for the "Big Four" or "Big Five" dominating the major industries.

Labor organizations, too, adopt restrictive practices. The common practice of limiting output or slowing the pace of work curtails the productivity of the economy. And the extension of labor organization and the prevalence of high wages tend to limit new enterprises whose initial success may require lower wages. Or the curtailment of profit because of high wages may prevent a small enterprise from accumulating funds for expansion. It is also significant that the large and well-established companies are somewhat protected from new competitors in this way.

These in summary are the principal business-labor-government policies and practices which increasingly place obstacles in the way of expansion,

so that the freedom and flexibility of the economy in the past are largely disappearing. According to the view under discussion, the consequences of this restriction were evident in the prolonged depression during the 1930's, and if allowed to remain they will impose chronically low levels of private investment in the future.

The Remedies

In general, those who hold the expansive economy thesis conclude from an analysis of the causes of depressed investment, output, and employment that there is the necessity for increased freedom of enterprise. This means freedom from "excessive" governmental intervention, from monopolistic practices by large corporate groups, and from union encroachment on the incentives and opportunities of enterprise to expand. It means a restoration of the traditional incentives of enterprise by the reduction of taxation and by removing undue risks imposed on investment. It means that government gives sympathy and support to enterprise, so that there is created the confidence which is prerequisite to long-term expectations and investment.

The tenor of such proposals is, it might be said, that of liberal conservatism. It questions alike the nostalgia for the "obvious and simple system of natural liberty" of Adam Smith and the past as well as the contemporary confidence in government control and planning. In simple terms, the expansive economy proponents believe there is little wrong with the economic system that freedom and enterprise cannot repair. Large public spending, reduction of savings through taxes, and close governmental participation in the affairs of the economy—the measures of the mature economy thesis—are considered by them simply as deterrents to the progress of the economy.

CAPITALIST DISINTEGRATION THESIS

According to a third major point of view, which may be called the capitalist disintegration thesis,[9] capitalism inexorably results in developments in political power and public attitudes destructive to itself. Capitalism cannot survive, not because of economic failure but because the success of the system in the productive sphere undermines the social institutions which protect and nourish it. Deriving disintegration from success is a novel in-

[9] The leading proponent of this view is Prof. Joseph Schumpeter. His diagnosis of the system is given in *Capitalism, Socialism, and Democracy* (Harper, 1942), especially Chaps. 5-14.

terpretation of capitalism, and this is one of the features which distinguishes this thesis from the Marxian socialist view.

The present interest in this large matter is limited to the relation of those fundamental political and social changes which accompany capitalist success to future material progress—increased investment, output, employment, and the standards of living derived therefrom. The view under discussion asserts that economic progress over the past two or three centuries is the outgrowth of the private enterprise system, or capitalism. Yet if self-destructive tendencies harbored within capitalism are impelling a gradual disintegration of the system, what are the prospects of future economic expansion? The inevitable result of these self-destructive tendencies is socialism of some sort, and it may be presumed that a fully developed socialist system will meet the problem of economic expansion in its own way.[10] However, the process of disintegration is believed to be gradual. Before it culminates eventually in socialism (if only temporarily) the capitalist system will probably have lost its capacity to assure high levels of investment, national output, and employment. In the view of some observers this stage is at hand, or not far off, for the process of disintegration is now believed to be very clearly defined—and irreversible. Relatively full employment in the next few decades is held to be unlikely.

Whether these prospects are immediate or at some distance is not vital to this thesis, for it encompasses the long term. The growth of capitalist society has been attended by political and social forces inimical to its long-term existence. Political groups which are either anticapitalist or merely indifferent to the system, have become more powerful. A generally hostile attitude toward enterprise thus comes to dominate legislation, the administration of law, and judicial interpretations. In this atmosphere, enterprise is unable to function as it had earlier; and consequently capitalism undergoes disintegration.

In summarizing this thesis it is necessary to examine its analysis of the nature of capitalism and the function of the enterpriser, and from this to proceed to the influences which are undermining the system.

Historical Capitalist Development

The history of capitalism as a process consists of periodic revolutions or economic upheavals, followed by long periods during which the effects of revolutionary changes have been absorbed by the economy. These upheavals reshape the structure of the economy; during the course of the

[10] See Chap. 42 for a discussion of the socialist economy.

process existing industries and capital values may be destroyed and unemployment may be widespread, but out of the disturbances emerges a revitalized and more powerful economy. These upheavals may be caused by the introduction of: (1) new methods of production, such as the use of steam or electricity; (2) new products, such as automobiles or electrical appliances; (3) new forms of business organization, such as the corporation; (4) new sources of raw materials, such as iron-ore deposits or oil; (5) new types of transport and communications, such as railroads, airplanes, or radio; (6) new markets and new selling techniques. As new factors are developed whole industries may be reduced, or even made obsolete, by new ones—the stagecoach gave way to the railroad—which may revitalize, transform, or create other industries.

Great changes of this sort are held, by this point of view, to characterize capitalism; they constitute the fact of capitalist progress. The similarity between this view and the explanation of economic progress ascribed to the first two theses is apparent, but there is this difference: in the present view capitalism can progress in no other way except by incessant revolutionary changes, and therefore only by the attendant destruction of the old economic structure and perhaps instability and unemployment.

The function of the enterpriser, then, in this ever-changing process is to "get things done." It is the enterpriser who reforms and revolutionizes the pattern of production by developing the industries, new commodities, new sources of materials, new markets and types of business organization. In so doing large-scale investment is undertaken and the system is propelled to higher levels of economic activity. To function, the enterpriser must be relatively unfettered—free from government controls, from the restrictions of unions and monopolistic practices, and from the curbs of "social consciousness" on his own part. He must be willing to challenge the economic and social status quo—to disrupt it by imposing revolutionary changes on the structure of the economy. The great achievement of the enterpriser and the capitalist system is making goods and services available to the masses and thereby increasing material welfare. Private enterprise has achieved more for the masses than was possible in any previous period.

The Growth of Hostility

Inherent in the development of capitalism, however, are the forces and changes bringing about the destruction of the system, the alienation of supporting and protecting groups, a weakening of the enterprise spirit and of the function of the enterpriser.

A critical and often hostile attitude toward private enterprise characterizes a growing segment of the public. In face of the obvious material progress of the system, why is this so? Any single explanation would be incomplete yet the salient points can be summarized. The *critical outlook* is itself an indigenous aspect of capitalism.[11] That is, capitalism imposes a *rationality* on man in his economic pursuits; this is formalized by price-cost-profit calculations. This rational outlook or mentality extends to all aspects of life, so that all social values and traditions tend to be subjected to the questioning, matter-of-fact attitude. The rational attitude is essentially critical; it tends to dissolve or erode traditional values and authority. Historically, the rise of the rationalist mentality, one aspect of which is the scientific outlook, served capitalism well, for the early enterprisers of the capitalist era were arrayed against the forces of feudalism—the Church and the landed nobility. To grow, capitalism had to combat the authority of the Church and nobility, and the social institutions and values which buttressed them. What basis in reason had the Church and the nobility for their enormous authority? The capitalist mentality was of necessity critical, for to establish its position it had to undermine existing authority and anti-capitalist values and traditions. This hardly does justice to the great issues involved, but the point developed stresses the essentially rational, matter-of-fact, critical mentality or outlook inherent in capitalism as a social process.

The proponents of the view under discussion assert that the critical outlook fostered by early capitalism has grown with it and increasingly turns against capitalism itself. Social criticism is no respecter of any institution. The forces loosed on feudalism have been attacking or questioning capitalism itself—particularly since the middle of the nineteenth century.

The spearhead of this critical revolt is the "intellectual"—a vaguely defined character—whose allegiance to existing authority and values is dubious at best.[12] But the defection of the so-called intellectual in itself may be more important as a symptom than as a real threat to the system. It is only when the dissident intellectual allies himself with other groups of actual or potential political power, groups which have real or fancied grievances against capitalism, that he may help to undermine capitalism.

This leads to the next point, namely, the defection or indifference of other numerically important groups which supported capitalism in its rise

[11] See Schumpeter, *op. cit.*, Chap. 11, for a discussion of this matter.

[12] It should be noted that "the" intellectual in Professor Schumpeter's interpretation does not include all intellectuals—he and innumerable others do not offend as does "the" intellectual in the drama of capitalist disintegration.

to power. The bulwark of early capitalism—farmers, retailers, and small shop owners—comprises some of these groups. However, the growth of big business capitalism has placed these groups in an economically defenseless position. "Big business" is constantly a threat to their welfare and often even to their existence. Consequently, their support of large companies and corporations is diluted by both fear and envy; and from them come few protests when government cracks down on big business. Not only is their emotional support of big business capitalism weakening, but the group itself is declining in relative importance. Farming at one time embraced more than three fourths of the population of this country, now less than one fourth. Small business proprietors no longer exert the influence of earlier times. Thus the bulwark of early capitalism—farmers and small proprietors—makes at best a weak ally of big business capitalism today.

Along with this, and partly explaining it, is the growth of urban population. Urban workers have relatively weak emotional ties with capitalism. Typically employees, whether "white-collar" or manual workers, their livelihoods are dependent on business organizations which become larger and more impersonal. Foremen, supervisors, and clerks in a personnel office inspire little or no personal support for the system, for frequently they lack it themselves. The bulk of urban employees have no prospect of being anything but employees, and are never rid of the fear of insecurity which is inherent in a system based on "risks."

In labor unions the indifference of employees to the system may be transformed into hostility, for those who are hostile find the union a means of expression. Furthermore, union leaders frequently enhance their own importance in the union bureaucracy by fomenting hostility. While unionism in this country has had a procapitalist tradition, its support of free enterprise tends to weaken as time goes on. Although urban employees are by no means in seething opposition to the system, their opposition tends to grow, and it is made articulate and effective by organization.

The urban population is naturally diverse and includes some groups that in other times strongly supported capitalism. Among these are the small proprietor and professional groups, who in large cities are but small minorities, with less influence on public opinion than in earlier times. Many others are linked to capitalism by virtue of their ownership of corporate stock. But here again the development of capitalism has tended to weaken the support of stockholders, i.e., the owners of corporations. Stock in a corporation, for practical purposes, is only a piece of paper which may be readily sold by the owner. The stockholder has relatively little real interest in the corporation, for usually he has no voice in its affairs. To him

the piece of paper often represents nothing more than a chance to make a speculative gain by selling the stock to someone else similarly motivated. By and large, stockholders have at most an attenuated attachment to the enterprise and vague misgivings about the management.

The majority of the urban masses, with their characteristic indifference to the traditional values of capitalism, are thus often attracted to the skeptical and critical opinions opposing the capitalist ideals and programs. Who are the spokesmen for such opinions? Largely, they are those intellectuals who have drawn on ideas and have evolved programs which promote radical changes in economic organization, and who exercise expression as radio commentators, columnists, writers who contribute to the "journals of opinion," and as experts and advisers in government who influence both public policy and publicity in the broad sense. These people frequently have a "public," largely in urban areas, having no deep emotional attachment to the traditions and values of capitalism, which is consequently susceptible to the critical views of the "intellectual."

Changes in Capitalist Functions and Mentality

The intellectual and social processes that create an atmosphere of indifference and hostility to capitalism as such also produce destructive tendencies within the enterpriser group itself. For one thing, the initial function of the enterpriser tends to erode as a result of the continued expansion of the system. It has been said that this function was essentially "getting things done," i.e., developing new methods, processes, and commodities and forms of business organization which propel and revolutionize the economy. The growth of corporate big business tends to destroy this function, because things get done, changes are initiated, chiefly by bureaucratized corporate management rather than individual enterprisers. The actual head of the business is increasingly a salaried manager, a careerist whose function is largely administrative. In the dominating field of big business, investment decisions thus are delegated to the hired administrator, restricting the traditional enterpriser or capitalist to the less profitable fields encompassed by small business.[13]

While the importance of administration in large corporations is undoubted, it should be understood that such administration is not necessarily capitalist. Administrators of bureaucratic organizations have much in com-

[13] This tendency is not only found within the operation of businesses but also in the field of large investment. "Captains of industry" and great bankers stand out less and less; figures like Jay Gould, Rockefeller, Morgan, Ford, and Carnegie are rarely found today.

mon, whether they administer a privately owned corporation or a governmentally owned corporation and other governmental agencies. And since the capitalist function is becoming depersonalized, bureaucratic, and therefore in a sense mechanical, the enterpriser in his traditional role is losing importance. Enterprise has achieved so much that it renders itself increasingly superfluous.

The emergence of the salaried manager as the custodian of capitalist functions, traditions, and values, together with the growth of anticapitalist criticism, has left its mark on the mentality or outlook of the capitalist individual himself. Receptive to public opinion, often as interpreted by his public relations counselor, the capitalist more and more compromises with the forces opposed to capitalism. Lacking strong emotional support from the public and perhaps the strength of conviction of earlier capitalists, he accedes to the trends toward "social welfare," which are not necessarily inimical to his purposes per se but which foster political regulation of enterprise. With Thomas Hobbes he seems to believe that

> riches joined with liberality is Power, because it procureth friends and servants: without liberality not so, because in this case they [riches] defend not but expose men to envy as a prey.

Thus the vigorous, effective, and ruthless spirit has gone out of the typical big business executive. But to what purpose? And with what consequences? His concessions to the hostile forces have not diminished their hostility. Each concession breeds new demands, for the politically vulnerable position of big business makes resistance to these demands precarious. The capitalist backs down before the growing strength of the organized worker, of the social reformer, and the public bureaucrat. While this may earn the capitalist the reputation of being public spirited, progressive, liberal, even high-minded, it saps his strength as a capitalist, and the virility of the system which created him. Schumpeter has described the capitalist position thus:

> The bourgeois fortress thus becomes politically defenseless. Defenseless fortresses invite aggression especially if there is rich booty in them. Aggressors will work themselves up into a state of rationalizing hostility—aggressors always do. No doubt it is possible, for a time, to buy them off. But this last resource fails as soon as they discover that they can have all.[14]

[14] *Op. cit.*, p. 143.

The Political Position of Capitalism

It is evident that the development of capitalism and its present position as thus interpreted are related to significant political trends. Political power in the United States has shifted from rural to large urban areas. Some elements in the urban population are openly hostile, many others are relatively indifferent to capitalist traditions and values. Can capitalism convert this indifference to its support? This is difficult because the capitalist lacks political insight and appeal—by training and disposition he is not well suited to political leadership. What is the capitalist political program? It is based on a belief in freedom of enterprise, which enables industry to continue its long-term expansion with long-term advantages to the masses in the form of jobs, income, and a higher standard of living. But the masses are not interested in the "long run," for as Lord Keynes once remarked, "in the long run we shall be dead." The urban groups, especially labor, are interested in the short term, and since enterprise in the short term is inherently unstable and insecure, they look to political leadership that promises them high wages, job stability, and economic security "here and now, for all."

Of recent times, therefore, many governmental policies are formed and much legislation passed which are inimical to the private enterprise system. Taxation is a case in point. The tax needs of government have increased steadily, necessitating a wide distribution of the burden. The principle of income taxation "in accordance with ability to pay" is politically popular with the masses, who favor high tax rates on corporation incomes and large individual incomes. Since these taxes lessen the incentives to invest, the private enterprise system is in essence undermined. Other legislative measures are similar in effect: wage and hour laws; encouragement of union organization; social security; regulation of business financing and business organization; restriction of certain competitive practices. The administration of such legislative measures and the courts' interpretation of them have followed the bent of the times and have been generally unfavorable to the free operation of capitalism.

The Economic Consequences

The consequences of capitalist disintegration on investment, output, and employment are evident. Investment is restricted and becomes increasingly inadequate. It is not held that the economy will soon sink into permanent de-

pression, for the analysis under discussion is predicated on the long period. Prosperity may prevail under capitalism for many years to come, according to this view, but the trends briefly indicated here will in time weaken the system and eventually destroy it. The last stages of disintegration, before some new order is established, will be the period of prolonged crisis. The social atmosphere is now, especially because of world conditions, sufficiently anti-capitalist, or merely indifferent, to weaken the incentives and limit the productivity of the system, but not sufficiently anti-capitalist to bring about an effective revolution to a socialist or other kind of order. But it is held, in the disintegration thesis, there will be a worsening of profit expectations as incentives erode, and with this will come chronically depressed levels of investment, national output, and employment.

The Futility of Remedy

Unrelieved pessimism is obviously the resultant outlook of this thesis. Nothing can be done to stem the disintegration of the private enterprise system. At best, some palliatives may ease the condition of the patient, but no cures can be expected. The lifting of restrictions on private enterprise, in the way proposed by the expansive economy thesis, may restore sufficient vigor to the system to bring about another long cycle of expansion. But the prospects of lifting restrictions for an extended period of time are not encouraging, for the political strength of the capitalist ideology is weakened progressively by both external and internal forces.

The remedies proposed by the mature economy thesis are rejected flatly, because they are regarded as part and parcel of the political influences which are destroying the system. A thesis that proposes to reduce the propensity to save by heavy taxation and thus weakens further the incentives to invest, and that seeks to increase the propensity to consume, is naturally inimical. The expansive economy thesis, on the other hand, is not inimical to the capitalist ideal, but it is regarded as conservative and illusionary from the disintegration point of view—optimistic without cause. While the first considers high taxes and other measures of the maturity thesis as deterrents to an otherwise healthy economy, the second regards such measures as inevitable symptoms of capitalist disintegration.

ECONOMIC PROGRESS AND STABILITY

As for the prospects of economic progress and stability, two of the three theses described here hold that they can be favorable. The measures pro-

posed to expand and stabilize the economy may be examined more specifically.

The Mature Economy View

The mature economy view, as already noted, proposes governmental spending (public investment) to supplement the relatively limited amounts of investment arising from the private sector of the economy. Public investment is considered a means of enabling private enterprise to continue in its own fields; and the combined investment of government and enterprise is deigned sufficient to sustain relatively full employment. This is not expected to impair private enterprise; indeed, if the private enterprise economy is not bolstered up by government spending, it will probably not survive in the long run.

In addition to assuring high levels of investment, national output, and employment, this policy of governmental expenditure is expected to achieve *economic stabilization* as well. Economic stabilization in this sense is taken to mean *an orderly expansion of employment year to year in keeping with the annual growth in labor supply.* How is this to be achieved within the framework of the mature economy thesis? In general by (1) controlled governmental spending, and (2) by increasing the consumption ratio or propensity to consume. These measures warrant examination in terms of economic stabilization.

In Chapter 6 it was shown that private investment is subject to rapid changes or fluctuations, and that these ups-and-downs are accompanied by ups-and-downs of national output, income, and employment. Indeed, it is frequently said that cycles of business activity are in actuality *investment cycles.* From this it follows that if the volume of investment were made stable over a period of time, national output would be stabilized too. If 28 billion dollars of investment is needed to attain 190 billion dollars of national output and relatively full employment, then stable investment is required for stability of output and employment.

How can investment be stabilized? Suppose that over a period of time private investment would average 23 billion dollars a year, and additional public investment, as a supplement, would average 5 billion yearly. But in the normal course private investment cannot be expected to be 23 billion dollars each year; in some years it might be 15 billion, in others 28 billion dollars. A policy of controlled public investment would therefore vary the amount government spent each year so that public and private investment

would total 28 billion dollars.[15] Securing just the right amount of public investment is the central point in the stabilization proposals of the mature economy thesis.

Coupled with this, however, are the various measures proposed to effect an increase in the propensity to consume—or decrease in the propensity to save. Among such measures are high wages, a widespread and adequate social security program, high tax rates on large incomes, low tax rates on small incomes, and job security to assure continuous income to wage earners. By increasing the propensity to consume an increased share of national output would take the form of consumer goods and services (i.e., the consumption-output ratio would increase). Since it is true that the output of consumer goods and services is relatively more stable than the output of capital facilities, national output would tend to grow more stable as the consumption-output ratio rises over a period of time. Or conversely, since capital-goods production is highly unstable, the smaller the share of gross output in this form, the less unstable would output be.

The mature economy thesis is more concerned about economic stability than the expansion or disintegration theses. Linked with its interest in stability is its preoccupation with *economic security,* i.e., the assurance of income (by the government) to those whose earnings have been cut off by unemployment, old age, industrial accidents, or other hazards.[16] Economic security through social insurance and related programs is in part regarded as a means of redistributing national income payments; it increases the propensity to consume and thereby contributes to economic stability.[17] And since stability is regarded as orderly expansion, there is perceived a comprehensive theory and program of (1) *economic expansion,* (2) *economic stability,* and (3) *economic security,* all related by the conception of the mature economy.

[15] The total required, of course, would increase gradually over the years as more investment is needed to provide jobs for the growth in labor supply and increased consumption for a larger population.

[16] The need for governmental programs to aid these groups is heightened by the factors underlying the mature economy analysis. A stable population has a growing number of aged persons requiring an increasing amount of old-age insurance; and industrial maturity makes the wage earner insecure, i.e., it threatens long periods of unemployment and therefore necessitates unemployment insurance.

[17] Chapters 39 and 40 discuss the major economic (or social) security programs, as well as the other means of affecting the distribution of income mentioned in this and the preceding paragraph.

The Expansive Economy View of Stability

This view is less concerned with stability than the view just summarized, or perhaps it is more accurate to say that the expansive economy proponents regard instability as a problem of only moderate dimensions in an expanding economy. Depressions occur but they are short-lived; indeed, they allow the economy to iron out maladjustments and therefore enable it subsequently to reach higher levels of output. Some attention is devoted to methods of *reducing instability,* particularly the control of the supply of money.[18] In the basic analysis in this view, it will be remembered, the problem of instability is regarded as an inevitable consequence of the deterrents imposed by government or private monopoly on enterprise. Although deterrents alone do not cause economic instability, for instability is inherent in the dynamic processes of the system, the deterrents undoubtedly aggravate an unstable condition. Thus in the absence of deterrents, economic *expansion* would be assured and *instability* would be of secondary importance only. Also, deterrents are regarded as the chief cause of widespread unemployment and therefore of *insecurity.* If deterrents were removed adequate incomes would be assured by employment in private industry. This would not totally remove the need for unemployment insurance or old-age insurance, but relatively full employment in an expanding economy would minimize the need for these programs.

This view, too, links together the great issues of economic expansion, stability, and security by relating each to its conception of a virile, growing economy. Stability is viewed as the absence of excessive booms and recessions, a condition made possible, not by an orderly expansion, but by a *dynamic expansion* which produces greater material welfare. From this dynamic expansion comes the best guarantee of employment and income and therefore security.

Stability and the Disintegration View

From what has been said of this view already, it is evident that economic stability has little or nothing to commend it. Since capitalism is a process of revolutionary economic change it is inherently unstable. However, this instability may be aggravated by unfortunate policies of government, and by the dislocations occasioned by war. Both phenomena involve a weakening of the capitalist system. Once weakened, governmental poli-

[18] See Chap. 11.

cies designed to enforce stability, assure security, and induce expansion only weaken the system further and hasten its transformation to a socialist or other type of economy.

CONCLUSION

The issues discussed in this chapter are of the utmost importance to the student of economic affairs, and to others as well, since from the description it is obvious that these interpretations of the economy bear directly on all persons' social and political philosophy. The views summarized above give different interpretations of the characteristics of the economy, chiefly as different conceptions of the nature and progress of the wealth of the nation. Each portrays the decisive role of investment: the incentives and opportunities, and the amounts of investment. The importance of investment, as a source of employment and income and as the demand for savings, has been analyzed in the preceding chapters in relation to the other major sources of employment and income, namely, consumption and governmental expenditures. This earlier analysis of the levels of national output, income, and employment may be fitted into each of the divergent views discussed above, for in each there is an interpretation of long-term investment prospects as well as of future levels of national output, income, and employment.

These divergent views emphasize another point, namely, that economic progress is governed by fundamental economic, political, and social forces. Population, geographic and technological factors, the sociological environment, and especially the attitudes or outlook of the people, the types of incentives, and the role of governmental policy—all influence the performance of the economy. This performance is the national output. The technical matter presented in the earlier chapters—the propensities to save and to consume, the relationships between investment, consumption, and governmental spending, the multiplier and acceleration principles—provides an analysis of this performance and the conditions necessary to obtain maximum output. But to understand developments of the past and potentialities of the future, it is necessary to study the fundamental economic, political, and social influences governing the growth of wealth.

PART III

THE MONETARY PROCESS

CHAPTER 8

The Nature and Functions of Money

The discussion of national output and national income in the preceding chapters took for granted a highly organized monetary process. National output is measured in terms of money: Money incomes are principal costs of production, and the exchange of commodities and services in the course of production and consumption is largely facilitated by money. The nature and functions of money, the types of money, monetary standards, banking facilities, and the conditions regulating the supply of money are matters of prime importance to an economic system. The performance of an economic system is dependent in many ways on its monetary arrangement. Expansion is predicated on a well-ordered system, and disruptions are caused by an ill-ordered one.

The money system of an economy can be viewed as both a mechanism for the facilitation of the exchange of commodities and a formula for economic measurement. Thus money is in a sense an instrument simply or merely useful for the functioning of the economy. But the money system of itself can cause disturbances and raise problems independent of its simple function as an instrumentality. To illustrate: the physical handling in shipping and storage of many commodities is facilitated by placing them in boxes. Hence the boxes serve a function of making easier the handling of the goods, and also can be used for the convenient inventory as well as measurement of the commodities. But suppose that the boxes, quite apart from anyone's wishes or plans, should appear in too large or too small numbers or should grow or shrink in size. It is obvious much dis-

turbance would result, particularly if the whole system of handling the commodities indispensably required the use of such boxes. Money sometimes behaves like these hypothetical boxes, with much the same disturbing consequences. In the periods after the two world wars the entire economic life of many countries was utterly disrupted when monetary mechanisms were impaired or escaped control. It has been said with some truth that bad money has caused more hardship and suffering than bad kings.

THE FUNCTIONS OF MONEY

The functions of money in the modern economy are (1) to act as a medium of exchange, (2) to be a common denominator of values, and (3) to form an economic link between the present and the past and future.

As a medium of exchange, money makes possible the highly complex system of exchanges which characterizes an economy of specialization. As it was noted in Chapter 1, specialization involves the production and distribution of many varied commodities and services which cannot be exchanged between producers and consumers directly. This is a commonplace explanation, since everyone is thoroughly accustomed to the use of money. To appreciate fully money's function as a medium of exchange, consider the problem of exchange in the absence of money. *Barter is the process of exchange when no single medium of exchange is used.* Even nowadays barter is occasionally resorted to, as for example when children trade playthings, or when a farmer trades pigs for calves. If no money of any kind is used, such direct exchanges can be based on open or tacit measurements in money. When two farmers trade three $10 pigs for a $30 calf, the transaction between them is greatly simplified by the fact that they agree on the money value of the animals. Although money is not passed between the swapping farmers, it is used by them as a common denominator; the value of the livestock is measured in terms of the common money unit. When a schoolboy trades a $10,000 pup for two $5,000 cats, he too is using money as a common denominator, even if his actual figures are inflated.

On the other hand, consider the process of barter in which money is not used even as a common denominator. Each individual is necessarily confronted with the problem of finding someone who has what he wants and wants what he has. Each individual with surplus products to barter for the surplus goods of others must find another individual whose wants and possessions will complement his. A man with apples who wants to trade them for pears must find another man who has pears and who wants to trade them for apples; thereupon, they must fix the terms of trade. How

many bushels or pounds of apples are worth how many bushels or pounds of pears? The value of apples in terms of pears, or of pears in terms of apples, will depend on the amount each trader has, the intensity of his desire to obtain the other commodity, and upon his skill as a bargainer. If the apple-owner wishes to acquire a great many different goods by barter, he must repeat this process at length, each time agreeing on a ratio of exchange once he has found a person willing to barter.

Barter is clearly an awkward method of exchange even where circumstances are simple. A further difficulty is concerned with units. Suppose someone who has a cow to trade wants salt. It is determined that a cow is worth five hundred pounds of salt, but the man with the cow wants only a few pounds. How can he arrange his barter transaction? He might perhaps barter the cow for several other commodities, exchange these in turn for still others, until finally he had some one thing which he could trade for the amount of salt he wanted, but because of this he may be burdened with several commodities which he must rebarter in order to get what he really wants. Therefore, because a cow and salt are available in quite different units, an exchange of one for the other may involve an exceedingly complicated series of barter transactions. In a money economy, on the other hand, such an exchange is simple: the cow is sold for $100, of which 10 cents is used to purchase salt.

Barter involves still another difficulty: the time factor in exchange. Many exchanges entail the delivery of goods at present in return for goods to be received in the future. An individual sells his services now, receives money, but holds the money until some future time when it is converted into other goods and services. But in a community using only barter as a method of exchange, it would be difficult to link the present and the future, because of the perishability of most commodities. Some commodities, of course, can be readily stored, and an individual who produces these could hold them for a period of time before bartering them. But for other individuals, in particular those producing services, this is impossible, or impractical. However, a barter system could make possible arrangements under which, say, a blacksmith could promise to provide his services to another person for a period of two years in return for which the second person would at the end of two years put a new roof on the blacksmith's house. Arrangements like this did indeed prevail widely in the Middle Ages, when barter was common since the use of money was very much limited. Thus although it is possible to exchange present goods for future goods under a barter system, the range of possibilities is narrow and the arrangements are necessarily clumsy.

In a money economy, all commodities are exchangeable for one another, almost instantly as it were, through the use of money as a medium of exchange. All commodities are instantly comparable in value with one another, because in the money system each commodity and service has an exchange rate, or price, expressed in the common denominator of money.

THE PECUNIARY SOCIETY

Modern society is often referred to as "the pecuniary society." The implication usually intended is that a money-using society tends to allow money to become the master of man, rather than remain his servant. Money, it is argued, dominates the minds of men in frequently perverse ways.

Consider some of the curious facets of money in modern society. Money, it has been noted, is a common denominator; the values of all commodities and services are expressed in money, as so many dollars and cents. There are many persons who habitually think almost solely in terms of money values, and to whom market price is far more important than, for instance, artistic merit. The casual visitor in an art gallery views the painting by an old master and asks, "How much is it worth?" If he is told that the painting is worth $100,000, the visitor is impressed. The same person sees a building the beauty of which is noted, and immediately asks, "How much did it cost?" He hears of a famous pianist, and again demands, "How much does this pianist make?" Similarly, rare books, scenic waterfalls, physicians —in fact anything and everybody—can be rated according to money value or money earnings. A hundred-thousand-dollar painting is twice as good as a fifty-thousand-dollar one, a building which cost 4 million dollars is twice as impressive as one which cost 2 million dollars, an artist or physician with twice the income of another is twice as skillful, ad infinitum.

This habit of thought is one of the aspects of the pecuniary economy, and it involves, as do all human activities, a question of standards. What standards should be applied? In the appraisal of a painting, the art critic employs an esthetic judgment founded upon his knowledge of the art and his developed taste; a dealer who may likewise have a sophisticated taste must also know and be able to calculate market prices; a wealthy arts patron perhaps has a developed taste, but he may also buy paintings simply to impress others; and again, there is the casual visitor to the gallery whose chief standard may be money value. Actually, the last is the simplest standard, for it can be applied to anything. Money value is a numerical standard: it resolves all questions of more or less, of greater or smaller.

Another characteristic is that, in society composed of individuals governed by laws, money represents power. Money is *purchasing power,* a means of commanding commodities. The possession or right to money is a means by which the individual makes a free choice among commodities. Money also commands services, and it is in this sense that the owner of money has power over other men. The possessor of 10 cents has the "power" to "command" a bootblack to shine his shoes. A $10 tip to the headwaiter gets a good table at a night club. Ten thousand dollars adroitly spent may command important political favors.

In some societies, other methods are used to command the services of one's fellow men: favoritism and corruption may be accomplished in other ways than by offers of money. Lord Keynes said that

> . . . dangerous human proclivities can be canalised into comparatively harmless channels by the existence of opportunities for money-making and private wealth, which, if they cannot be satisfied in this way, may find their outlet in cruelty, the reckless pursuit of personal power and authority, and other forms of self-aggrandisement. It is better that a man should tyrannise over his bank balance than over his fellow-citizens; and whilst the former is sometimes denounced as being but a means to the latter, sometimes at least it is an alternative.[1]

WHAT MONEY IS

Up to this point there has been no need to define money precisely. Most people seem to regard money as the coins and bills they carry about with them. "The" money they have "in" the bank is counted along with the money in their pockets. But actually money in the pocket and money in the bank are two quite different things, although for some purposes no special distinction need be drawn between them. As another example of the everyday use of the term money, one may say, "Smith has lots of money." Smith may own bonds, mortgages, real estate, stores, and other valuable things, but conceivably he could be possessed of only $10 in coins and bills. He would still be a wealthy man. His wealth happens not to consist of money although it is measured in money.

In economic analysis money is currency and demand deposits. Both of these terms require explanation; however, the background of these terms must first be examined. As was noted, money functions as a medium of exchange, as a common denominator, and as a link between the past, present, and future. Anything which performs these functions could be called

[1] *Op. cit.,* p. 374.

money. More briefly, any means of payment passing at its nominal or face value is money.

A great many things are used as means of payments. A partial list[2] includes: gold, currency (i.e., coins and paper money), demand deposits in banks, time and savings deposits in banks, open lines of credit with merchants, government bonds. A foreign government may ship bars of gold to the United States and thus complete a financial transaction. Of course coins and dollar bills most commonly are used to make small payments. When somebody writes a check in payment, he is using a demand deposit in a bank. If a bank will allow its savings depositors to write checks to make payments, the savings (or time) deposits are also means of payment. An "open line of credit" is used as a means of payment whenever someone buys commodities under an established agreement on credit. Government bonds of the marketable types are sometimes transferred from one person to another for the purpose of making a payment and in lieu of a payment by check or in currency.

As a matter of fact, almost anything can be used to make payments. If anything and everything were used, a system of barter would be in effect; any commodity, being directly exchangeable for any other, would constitute means of exchange, or payment. The peculiar quality of money, however, lies in its being a *common* denominator for exchange. It follows therefore that the common denominator should consist of one thing or, at the most, of only a few things. This is indeed what is found in modern economies. While many things are used to make payments, by far the greatest bulk of all exchanges is effected through the transfer of ownership of currency and of bank deposits.

Money, then, can be simply defined as currency and demand deposits. This is the most commonly used definition. Whenever the word "money" is used in the following pages, it will mean only: currency and demand deposits.

Currency requires no special explanation at this point. It consists of coins and of paper bills, both of which are issued by the government or by government agencies.[3] A dollar bill is actually just a piece of paper, but what is it? It is a debt owed by the government to whomever owns the dollar bill. A dollar bill is an asset to its individual or corporate holders, but it is a *debt* of the government. A United States Savings Bond is an

[2] Adapted from Woodlief Thomas, "Money System of United States," *Banking Studies*, Board of Governors of the Federal Reserve System, 1941, p. 297.

[3] That is, Federal reserve banks, which are strictly speaking instrumentalities of the Federal government, although they are privately owned.

asset to its owner, but it is a debt of the United States. The difference between a savings bond and a dollar bill is that the bond cannot be used to make payments, while the dollar bill exists for that very purpose. Paper bills, then, are the circulating debts of the government.

Demand deposits are the circulating debts of commercial banks. This definition is not so obvious as the statement that dollar bills are circulating debts of government, since to be convinced that a dollar bill is a debt, it is only necessary to read the printing on a dollar bill. A demand deposit—or checking account—is a deposit which is payable by the bank on demand, payable at any time (during business hours) when payment is requested by the depositor. In contrast, a time deposit—or savings deposit or savings account—is legally payable only after a specified period, such as 30 days. That is to say, a bank has the right to delay payment on a time deposit. In normal practice, as an accommodation, banks do pay time depositors on demand. But both a demand deposit and a time deposit are debts of the bank and assets of the depositors. A bank deposit is not money "in" the bank; actually it is a liability or debt of the bank.

Suppose that John Smith decides to open a checking account, which is of course the same thing as a demand deposit, with the First National Bank. John Smith begins by depositing $100. This means, in most cases, that he transfers to the First National Bank $100 in currency, and in exchange he receives a passbook in which the bank acknowledges that Smith's deposit is $100. At the same time the bank adds Smith's name to its list of depositors and notes on its own ledger that his deposit is $100. John Smith walks out of the bank satisfied now that he "has a hundred dollars in the bank." He thinks of "his money" as being "in" the bank. If Smith a week later withdraws $50, he thinks of it as getting his money out of the bank, although of course he takes whatever currency the teller hands him and does not insist upon getting the same bills back.

Quite apart from what Smith thought about it, what actually did happen when he made the deposit? He entered into a contract with the bank, under which the bank accepted his currency and promised to return the same amount of currency at any time Smith demanded it. The bank owes him $100. Smith's passbook and the bank's ledger say nothing more: the bank owes Smith $100.

Suppose Smith writes a check for $20 to William Jones in payment of a debt. A check is a written order to a bank to pay a certain sum of money to the person named on the check. Jones could cash the check at the First National Bank, which pays Jones $20 in currency, and deducts $20 from Smith's deposit, i.e., from the sum owed him. If Jones also has an account

with the First National Bank, the bank adds $20 to Jones' account and subtracts the same amount from Smith's. If Jones' bank is the Second National Bank and he deposits Smith's check there, the Second National Bank credits Jones' account with $20 and presents the check to the First National Bank. The First National Bank pays the Second National Bank the $20 and debits (i.e., deducts from) Smith's account. The sum total of deposits in the two banks together remains unaltered, but the Second National owes Jones $20 more and the First National owes Smith $20 less.

The total demand deposits of a bank are what it owes to its depositors. The grand total of all demand deposits in the country as a whole is what all banks owe to all depositors. These deposits—debts of the banks—can be transferred from one person to another by means of checks, as indicated. The bulk of all payments in a money economy are in fact made by check.

The reason that checks are more commonly used than currency for larger payments is that they are far more convenient. Checks can be written in the exact amounts required for transactions, while currency often presents a problem of making change. A single piece of paper can be used for a payment which would require a bulky package of currency. Checks are legal receipts for payment, because when the drawer (writer) of a check gets it back from the bank in canceled form the payee's signature is on it. Checks are a convenient method of making payments by mail. Checks are safer; if stolen, payment on them can be stopped. Stolen currency can be recovered only rarely, with police assistance. Bank robbers try to seize currency, but never bother with checks. It is no doubt easier to forge checks than to counterfeit currency, but the sufferers from successful forgery are banks and others who cash checks, not the persons whose names are forged (unless these persons do not realize that their names are being forged).

What a Dollar Is

The subject of money has been discussed without specific attention to the question of what a dollar is. It is obvious that a dollar is a unit of money in the United States, just as a franc and a pound sterling are units of money in other countries. But what kind of a unit? And how can this unit be defined?

A dollar is a unit of liquid claim. If the amount of money in existence is 100 billion dollars, then $1 is one-hundred-billionth of the total of generally acceptable means of payment. The owner of a dollar bill can go forth to buy things, with one-hundred-billionth of the total of acceptable means of

payment. If, however, the total amount of money was 1 billion dollars, the holder's claim against things offered for sale would be one hundred times larger than in the former instance. The importance of a dollar as a unit of liquid (i.e., generally acceptable) claim depends, then, upon how many dollars are extant and how many other things are available to be bought with dollars. This relationship between dollars and things is a variable one. A dollar commanded more things in 1939 than it did in 1947; therefore, it is said that the purchasing power of a dollar was higher in the previous period. The dollar is not a constant unit like the ton, the yard, and the gallon. The claim a dollar gives to its possessor fluctuates.

Another definition of a dollar is that it is a lump of gold, nine tenths pure, which weighs $15\frac{5}{21}$ grains. This is the basic legal definition of a dollar in terms of gold, and of gold in terms of dollars. Among other things it means that if gold coins were minted, a dollar coin would weigh $15\frac{5}{21}$ grains of gold. Before examining this ratio, however, it is necessary to consider the whole question of gold in relation to money in the light of monetary standards.

MONETARY STANDARDS

Monetary standards are methods of defining the unit of money and of controlling the supply of money. Currency and demand deposits are money, yet only pieces of paper in actuality. What makes them valuable? What is the security behind these pieces of paper? This is fundamentally the question of the money system and of the monetary standard.

Under the *gold standard* as it formerly existed, all money was freely convertible into gold. A ten-dollar bill or a ten-dollar bank deposit could be converted into a ten-dollar gold coin. The individual had the right to demand at any time that he be paid gold in exchange for paper currency or for the amount of a bank deposit. Since all forms of paper currency, all coins, and all demand deposits in banks could be converted into gold, they were interchangeable with one another. Gold was the standard. The gold standard will be examined further in the discussion following, but for the moment it suffices to note that the gold standard represented two things: (1) the use of a physical commodity (gold) as the security for money, and (2) a system of monetary control. Consider now the use of a physical commodity as standard money.

In very general terms, money is anything which is used as means of payment. In the history of mankind a great variety of commodities have

served as money: salt, tobacco, cattle, iron, stones, beads, shells, copper, silver, and gold. In parts of Europe during 1945-47, cigarettes served as money. In fact almost anything can perform the functions of money in a simple economic society provided only that it is in wide demand and that it is physically fairly durable. The word "pecuniary" itself is a reminder that cattle (Latin, *pecus*) were once a common denominator of value. When cattle served as money they were a common denominator, and perhaps an occasional means of payment though only for large transactions. A man's wealth and status could be measured in terms of the cattle he possessed, even if small transactions had to be handled otherwise. From ancient times, however, and in nearly every type of society, gold and silver have occupied a dominant position as the physical means of payment and common denominator of value. Where, as in colonial Virginia, tobacco was used as money, the reason was the availability of tobacco and its acceptance in payment in a community where gold was lacking. Other commodities than gold and silver have been employed but rarely when these could be used.

Why the preference for gold and silver? These commodities are metals; they do not rust and are durable. No physical loss takes place when gold or silver are stored and even large quantities (in terms of value) occupy little space. Both metals can be cut up into small pieces, or small pieces can be melted and formed into large pieces, without impairment of physical attributes; both are divisible in this sense. Gold dust is as valuable as a lump of gold of the same weight. Gold and silver alike are readily recognizable; little training is required to enable anyone to identify them easily. In this respect gold and silver are superior to precious stones as widely acceptable means of payment, for gems are cognizable only to persons with special training. Furthermore, gold and silver are uniform. Both can be mixed with alloys to produce coins of any desired degree of hardness. All of these are reasons for the persistence of gold and silver as money for centuries. But besides these practical reasons there is a second factor accounting for the preference for gold and silver: their place in the myths and folklore of mankind, especially of Western man. For centuries the greed for gold has been the force behind much human effort, for the quest for gold has led to the exploring of continents, to the alchemical effort to transmute the baser metals, mass migrations, and, as in South America, even to the destruction of civilizations. There has indeed been something magic about gold, and there still is. Gold continues to be the means for final settlement of international payments, and remains a major factor in modern monetary systems.

The Gold Standard

Although gold and silver had been the principal forms of money for centuries, most of the countries of the world adopted the gold standard only as late as the nineteenth century. The major features of the gold standard are: (1) legal definition of the monetary unit as a certain weight of gold; (2) the free and unlimited coinage of gold; (3) convertibility of all coins, paper currency, and bank deposits into gold at a legally fixed rate; (4) freedom to import and export gold.

In the United States the gold standard may be said to have existed from 1879 to 1933. Before 1879 the monetary standard was somewhat confused. In 1792 Congress established a bimetallic standard, when Alexander Hamilton had recommended to Congress the adoption of the system then current in Europe. Under a bimetallic standard both gold and silver are standard money, both are coined freely. Free coinage means unlimited coinage: any person has the right to bring any amount of gold or silver to the government in exchange for an equivalent amount of the metal in coins. Thus, in 1792 Congress defined a dollar as either 24.75 fine grains of gold or 371.25 fine grains of silver.[4] It was expected that both gold and silver would be brought to the mint, coined, and put into general use. The expectation was not fulfilled, however, and the few coins minted were nearly all of silver. The ratio—called the mint ratio—of the legal weight of a silver dollar to that of a gold dollar was set at 15 to 1. That is, a dollar was declared to be a definite lump of gold, or a lump of silver 15 times as heavy. The assumption was that gold was 15 times as valuable, weight for weight, as silver. But besides the legal determination of the exchange value of gold and silver there was another ratio, the market ratio, which measured the values of gold and silver in terms of each other as reflected *in actual commercial transactions.* These two ratios had to be identical, or very nearly so, if both gold and silver were to circulate together as coins. If the mint ratio were 15 to 1, and the market ratio was 16 to 1, then it paid anyone having silver to take it to the mint to be coined, and it paid anyone having gold to sell in the commercial market. In fact, if a difference such as this persisted, the coinage would come to consist of silver only, and gold would disappear. Silver in such a case was said to be "overvalued" at the mint, while gold was "undervalued." An overvalued currency drives an under-

[4] "Fine" means pure, i.e., without alloy. The weights of gold and silver are measured by the troy system, in which 480 grains equal 1 ounce, and 12 ounces equal 1 pound troy. One pound troy equals 0.82286 pound avoirdupois, the usual pound.

valued currency out of circulation. This principle is "Gresham's law," which is not a statute, as careless students have often thought, but simply an economic platitude attributed to one of Queen Elizabeth's ministers.

Beyond the difficulty of maintaining these ratios, bimetallism did not operate well during the early part of the nineteenth century also because there were then in circulation fairly large quantities of worn and clipped foreign coins. These served the purposes of trade and so almost obviated the necessity for full-weight government coins. This is still a further illustration of Gresham's law. In 1834 and in 1837 Congress altered the legal weight of the gold dollar. It became 25.8 grains, nine-tenths fine, or 23.22 grains fine. This legal weight remained unchanged until 1934. The silver dollar continued at the same weight but the change in the weight of the gold dollar altered the relative positions of the two metals. Silver had been overvalued at the mint before 1834, and gold became overvalued after that date. The gold discoveries in California and Australia in the 1840's and 1850's added to this overvaluation of gold at the mint. Thus after the 1830's the coins turned out by the United States Mint were predominantly gold coins. The currency in circulation, on the other hand, consisted largely of paper currency issued by state banks. Since much of this paper currency was in very small denominations, gold coins were not widely circulated in ordinary business transactions.

In 1873 Congress passed an act revising and codifying the coinage laws. The silver dollar was dropped from the list of coins which could be minted on a free and unlimited basis, and this step was tantamount to the abandonment of bimetallism and the adoption of gold as the standard. Henceforth, and until 1933, only gold could be brought by anyone and in any amount to the mint to be coined. This was the "crime of '73," the end of the free and unlimited coinage of silver. Silver coins have of course been minted since then, but only according to the directions of Congress. The action of Congress in 1873 was simply a recognition of a situation which had long existed: the silver dollar had not been coined for a long period. It was after 1873 that silver began to be mined in large quantities and that agitation for restoration of free coinage of silver began. Consideration of this problem is undertaken in Chapter 11.

The gold standard was thus legally adopted in 1873, though it is doubtful whether this was actually intended. It was not until 1879, however, that the gold standard began to function, for only in that year did the government begin to redeem in gold the greenbacks issued during the War between the States. In 1900 Congress deliberately formalized the gold

standard by an act which obligated the Secretary of the Treasury to maintain all currency and coin on a par with gold.

So far the convertibility of money into gold has been stressed as the important feature of the gold standard. This is the traditional view and is readily understandable. The plain man of two or three generations ago knew how to test the content of a gold coin by biting it or ringing it on a table (or bar): and since paper money could be exchanged for gold, the paper was as good as gold. Anything as good as gold was good enough.

The less obvious but more important factor to be considered in the gold standard was its existence as a *system* of monetary management. Gold coins circulated, and all other forms of money had gold reserves behind them: gold was the standard. Behind some of the currency, namely gold certificates, the gold reserve was 100 per cent; behind other forms of currency and demand deposits of banks the gold reserves were partial. But the legal reserve requirements were changed very rarely. Hence, the total amount of money was directly related to the amount of gold in existence, i.e., there was a more or less fixed ratio of amounts established between gold and other money. No one controlled the amount of gold in existence. This depended chiefly upon (1) the accumulated stock of gold coins and bars which had come down through the ages, (2) the fluctuating volumes of imports, exports, and other foreign trade transactions, (3) the amount of gold mined each year, and (4) the amount of gold coming from and going into nonmonetary uses. Furthermore, changes in the amount of gold were not great from year to year or, for that matter, from decade to decade. Therefore, it seemed to be safe to tie the volume of money to the volume of gold. This system was followed not only in the United States but also in many other countries, and by the second half of the nineteenth century it had come to be regarded as the only sound system of controlling money. The control was automatic; that is, it was beyond conscious direction. All that a government had to do was to maintain convertibility of its money in gold by itself providing interchangeability of gold and money, and by requiring commercial banks to do the same thing.

There was much to be said, indeed, for the automatic-control aspect of the gold standard. Governments in the nineteenth century (and earlier) had caused financial confusion when they had suspended conversion of money into gold. Great Britain during the Napoleonic wars and the United States during the War between the States had both seen inflationary (i.e., price-raising) overissue of paper currency. There had been numerous other instances of governmental irresponsibility. Private banks had provoked disturbances by issuing paper currency without assuring adequate converti-

bility into gold. Hence the great monetary phobia of the nineteenth century was "overissue." Governments were considered to have a tendency to become financially irresponsible, because of the temptation to meet financial exigencies by issuing paper money. The method of meeting the problem was establishing a money which was always convertible into gold, i.e., by the gold standard.

There was the one other important aspect of the gold standard, i.e., international trade and finance. The international aspects of the gold standard are treated in detail in Chapter 35, but at this point it may be observed that where several countries are all on the gold standard, the rates of exchange among all of their monies are stable within very narrow limits and prices in the several countries are closely related. Thus the gold standard can also be viewed as a mechanism for controlling foreign exchange rates and, to some extent, of controlling the general levels of prices. Here too the gold standard operated as an automatic mechanism in effect and function though actually directed by no one.

At the beginning of World War I the gold standard was abandoned by the belligerent nations. Convertibility of money into gold was stopped, and restrictions were placed upon the export and import of gold. Although the United States continued to maintain convertibility, it was affected by other countries' regulation of the international traffic in gold. Although during the 1920's most countries returned to the gold standard, many abandoned it again during the early 1930's. The reasons for the abandonment of the gold standard after 1914 are complex. The functioning of gold in international finance was the crucial factor.[5] For the present introductory purposes it is perhaps sufficient to point again to the nature of the gold standard. Its automatic character fitted the pre-1914 era. Just as governments interfered very little with other economic processes they were content not to direct monetary functions. The period since 1914 has been marked by increased governmental regulation of many things, including money. In this period the idea and the practice of some sort of "managed" money triumphed over the old principle of the gold standard.

Other Monetary Standards

Besides the gold standard and the bimetallic standard, many other monetary systems have existed and still others have been proposed by advocates of reform. These latter include money systems based on wheat, or on

[5] See Chap. 35.

wheat and cotton, or on a list of staple commodities, or on a unit of energy (such as the "erg"). Such proposals for reform will be considered in Chapter 11; meanwhile, standards other than the gold which have actually existed will be examined.

An *inconvertible paper standard* is one in which the standard money consists of paper notes issues by a government. These paper notes or bills, exemplified by the greenbacks of the period during and for a while after the War between the States, are considered standard money since all other money such as bank notes and bank deposits are legally secondary. Inconvertible paper money has often been termed *fiat money*, circulating by virtue of the fiat or order of the government and not because of the value of the piece of money itself. A gold coin was valuable as a lump of gold, independent of the fact that a particular government had coined it; but a piece of paper can serve as money only if the people possessing or receiving it have confidence in the issuer of the paper. This confidence can be based on knowledge that the issuer can and will redeem the paper in something valuable in itself such as gold, or upon the faith that the issuer will some day redeem a stated promise to pay in gold. An additional factor is the holder's faith that the issuer will not issue too much currency and so diminish its value. The greenbacks were not redeemed in gold until 1879. In the period before 1879 they could be exchanged for gold, but only for gold owned by private traders. And they were exchanged at a discount: for instance, $1,000 face value of greenbacks exchanged for $800 gold. Or, inversely, gold was at a premium: $1,000 gold was worth $1,250 in greenbacks. The gold premium fluctuated, according to traders' opinions as to the future of the government of the North during the War between the States; the premium declined as the North's military fortunes were improved. In the period after 1865 the gold premium steadily fell as it was realized that the greenbacks would be redeemed in gold. As soon as they were in fact redeemable, the gold premium fell to zero: gold and greenbacks became freely exchangeable, one for the other at par.

Many other examples of the use of inconvertible paper currency may be found. Great Britain, for instance, had such a system during the period of the Napoleonic wars. Most countries during and shortly after World War I resorted to a system of inconvertible paper currency. An inconvertible paper standard has always been the object of severe criticism, the reason being that in contrast to the gold standard, a paper standard lacks a means of limiting the total supply of money. In a system in which all money must be converted into gold, the amount of money is controlled by the amount of gold, but when there is no requirement of convertibility, the paper cur-

rency systems may result in an overissue of the fiat paper currency. Overissue has always characterized the system's operation in the past. "Overissue" means that so much currency is issued that inflation (i.e., rising prices) results. Financial exigency of the issuing government is usually the cause of overissue. When governments in the past were unable to raise sufficient funds by taxation or by selling bonds, they paid expenses by issuing paper currency. Hence the amounts issued depended upon the financial position of the government, and not upon considerations as to how much money an economy ought to have.

The United States Monetary Standard Since 1934

The United States abandoned the gold standard in April, 1933. This step meant two things: gold could be exported from the country only after obtaining a license from the Treasury, and currency and bank deposits could no longer be redeemed in gold. Furthermore, all monetary gold and all gold certificates (a special form of currency backed up dollar for dollar in gold set up as special reserve fund) had to be turned in to the Treasury. Holders of gold coins, of gold bars, and of gold certificates were paid by Treasury check. Not only individuals and businesses, but also all banks were required to turn in gold and gold certificates. The gold of the Federal reserve banks—by far the largest part of the country's monetary gold supply—was transferred to the Treasury in exchange for gold certificates, which are now held only by Federal reserve banks. Since 1933, then, all monetary gold has therefore been in the possession of the United States Treasury. Nonmonetary gold, of course, continues to be in private hands, because gold is used for industrial purposes, such as for jewelry and dental work. Coin collections of numismatic value were also permitted to remain in private hands. Apart from this, however, gold coins have not circulated at all since 1933. In fact, they ceased to exist, for all were melted down and formed into bars, or ingots, of gold.[6]

After calling in all gold, the Treasury began in 1933 to buy gold at steadily higher prices. The legal price of gold since 1837 had been $20.67 per fine ounce. This is equivalent to 23.22 grains of fine gold in a dollar. Since there are 480 grains to an ounce, the proof of the equivalence of the two figures is a matter of arithmetic. In simple terms the relations can be

[6] The complete lack of gold coins in the United States proved embarrassing during the North African invasion of 1942, when it was found necessary to have gold coins in order to bribe potentially hostile groups in the region. Canadian gold coins had to be hurriedly procured.

thought of in this way: if an ounce of gold is $20, then $\frac{1}{20}$ of an ounce must be $1, and $\frac{1}{20}$ of an ounce is 24 grains. Again, if an ounce of gold is $30, then $1 is $\frac{1}{30}$ of an ounce, or 16 grains. Thus, the higher the price of an ounce of gold, the less the number of grains in a gold dollar: raising the price of gold results in reducing the physical weight of the gold dollar. Thus when the Federal government in the summer and autumn of 1933 continued to increase the price it paid for gold, its action was tantamount to decreasing the weight of the dollar, although this was not then the actual legal effect. By the end of 1933 the price of gold had been raised to $35 an ounce, the equivalent of $1 weighing $15\frac{5}{21}$ grains (nine-tenths fine).[7]

The Gold Reserve Act of 1934, approved on January 30, 1934, ratified the emergency measure of 1933 and established the present money system of the United States. The price of gold was thereby fixed at $35 an ounce, the weight of a gold dollar at $15\frac{5}{21}$ grains (nine-tenths fine). These definitions have been maintained since 1934. The President's authority to reduce the weight of the gold dollar to one half of the original weight of 23.22 grains (fine) was extended; but this power has not been used since 1934. The redemption of currency in gold was prohibited by the Act of 1934, except as the Secretary of the Treasury would permit it. The act provided also that the bulk of the "profit" realized by the Treasury in its gold operations of 1933 be set aside in a stabilization fund. This fund began with 2 billions of the profit, to be used for the purpose of carrying purchases and sales in the foreign exchange markets. The "profit" existed because the gold seized by the Treasury in 1933 was taken in at 23.22 grains to the dollar. On January 30, 1934, this gold was declared to be dollars at the rate of $15\frac{5}{21}$ grains. Therefore the same amount of gold became, at a stroke of the President's pen, more dollars (approximately 2.8 billion more).

The various types of currency now in existence, to be described shortly, are not redeemable or convertible into gold, nor are the demand deposits of banks. Yet gold is used in the system. Gold certificates are the ultimate reserves behind bank deposits and most of the currency. Gold certificates are exchangeable for gold—but only within the government, i.e., between the Treasury and the Federal Reserve System. Gold is still used in international financial transactions. Gold imports and exports still affect the volume of money, but in an indirect and complicated way.

The United States money system is therefore a system of paper currency and bank deposits which are inconvertible but which are part of a monetary

[7] For a discussion of the reasons for raising the price of gold in 1933 and the results of this action, see Chap. 11.

mechanism in which gold is used. The mechanism itself is a controlled one; it is a managed monetary mechanism. The control or management of the money mechanism is not, however, complete in all respects. Precise control of the total quantity of money, for example, has not yet been achieved. The system has been called a managed inconvertible gold bullion standard. This expression is awkward, but necessarily so, for many factors must be covered by the definition. The monetary system is "managed" by the Treasury and by the Federal Reserve System; gold *bullion*—bars and ingots—are used instead of gold coins; currency and demand deposits in banks are not convertible to gold, yet gold is used.

The word "system" has been stressed in this description of monetary arrangements in the United States. The former gold standard, it is clear, was an automatic mechanism. The present standard is more complicated and sophisticated, and is actually a managed system. The overwhelmingly powerful financial position of the national governments is an important factor in the rise of the present systems. In the past, governments were financially weak, having occasionally to ask private bankers for aid. Small wonder, then, that the paper money of governments was often subject to mistrust. Within the last generation, however, relations have changed to such an extent that the banks are dependent upon the government. Thus, for example, the national debt is held by the banks and insurance companies as well as by other corporations and by individuals. The holders of the national debt are the owners of government bonds. Government bonds—also pieces of paper—are assets to their owners. The most important asset of banks are government bonds, and banks are therefore dependent on the value of these bonds, i.e., upon the government's paying interest and principal. The holding of government bonds by banks is but one facet of the dominance of government over the whole financial apparatus which has so largely contributed to a change in national monetary systems.

Convertibility into gold is actually no longer important. Long before 1933 gold coins were curiosities to the ordinary man. Gold coins existed, but were found largely in vaults, not in hand-to-hand circulation. When the convertibility of currency and deposits was stopped in 1933, there were widespread fears at the time that something sinister would happen to the money of the individual and the businessman. But nothing directly observable did happen. Currency went on being used as before, and checks were written and taken in payment, as before; so far as individuals and businesses inside the country are concerned the working of the money system has not been affected. There has been no "gold premium" as there was during and

after the War between the States, since private trading in gold was prohibited; and there has been no black market in gold, because the Treasury's price for gold is one at which gold can be turned into money in any quantity. Furthermore, gold is available for industrial purposes and for export under license at $35 per ounce.

THE CURRENCY SYSTEM OF THE UNITED STATES

The currency system of the United States is a complicated one. For September 30, 1946, the types and value of currency in circulation are shown in Table 17.

Table 17 **CURRENCY IN CIRCULATION**

Gold certificates	$ 50 millions
Standard silver dollars	144
Silver certificates and Treasury notes of 1890	2,010
Subsidiary silver coin	861
Minor coin	325
United States notes	315
Federal reserve bank notes	451
National bank notes	112
Federal reserve notes	24,237
Total	**28,507**

Source: Federal Reserve Bulletin.

It is to be noted primarily that all currency is issued by, or on the responsibility of, the United States Treasury or the Federal reserve banks. Before 1865, paper currency was issued by state banks, which are banks chartered and regulated by state governments. From the time of the War between the States to 1935, paper currency was issued by national banks, which are chartered and regulated by the Federal government. Since 1914 the Federal reserve banks have issued paper currency, the Federal reserve note being in fact the principal constituent of the currency supply. Federal reserve banks are public institutions whose primary purpose is to regulate the entire banking system.[8]

Consider now the appearance of a Federal reserve note (any bill of five-dollar or higher denomination is likely to be a Federal reserve note). The name of the issuing Federal reserve bank appears, and also the statement

[8] See Chap. 10.

that the note is redeemable in lawful money at the United States Treasury or at any Federal reserve bank. This statement means two things: that Federal reserve notes are liabilities of both the Federal reserve banks and the Treasury, and that the note is payable in other paper currency only, for "lawful money" is currency issued by the Treasury. Thus, one five-dollar bill is redeemable in another, and the possessor of it may derive what consolation he can from that fact. The Federal reserve note also says that it is "legal tender for all debts, public and private." Formerly the question of the legal-tender powers of a particular form of currency was an important one, because by law some forms of currency were assigned only limited legal-tender powers. Legal tender is money which a creditor may not refuse in payment offered to him; should he refuse legal tender he has no recourse at law against the debtor for payment. Legislation in 1933 made all currency legal tender.

Behind the Federal reserve notes there are reserves. These consist of gold certificates, to the extent of at least 25 per cent of the value of the Federal reserve notes, with the remainder consisting of government bonds or of certain short-term obligations of businessmen, such as *commercial paper.* Commercial paper consists of notes signed by businessmen in exchange for loans from their banks. Such paper may be used as collateral for Federal reserve notes when these banks rediscount (sell) it or pledge it at the Federal reserve banks to obtain funds. From 1913 to 1932 the security behind the Federal reserve note was a minimum of 40 per cent in gold, the remainder in commercial paper only. Government bonds were introduced as a security for Federal reserve notes in 1932.

The original intention of Congress was that by requiring both gold and commercial paper, the Federal reserve note would be safe because of the relatively high gold backing, and further that it would be responsive to the needs of business by being backed also by commercial paper. It was believed that the volume of commercial paper would fluctuate with general business conditions, that therefore the backing for these notes would be available in larger or smaller amounts (as businessmen increased and decreased their borrowings from banks), and that in this way the volume of Federal reserve notes would be made to fluctuate automatically with the level of general business conditions. In 1932, however, business borrowing from banks had fallen so low that there was little commercial paper in existence; hence the permission in this year to substitute government bonds.

National bank notes are still in circulation, although they have been in

the process of retirement since 1935. They were backed up dollar for dollar by government bonds. That there are still some notes in circulation means only that not all have been returned to the Treasury; presumably some have been lost, or are otherwise unaccounted for.

Small amounts of gold certificates and Treasury notes of 1890 are listed as being in circulation, but these too must be considered unaccounted for. Gold certificates, backed up fully by gold, were called in in 1933. They are not now in actual circulation, but are used only between the Treasury and the Federal reserve banks. The Treasury notes were originally issued in 1890 with a silver backing, but these early notes have long since been called in.

Federal reserve bank notes, also called national currency, were issued in small amounts after 1913. During World War II considerable quantities of them were released from a printing originally made in 1933, when it was believed necessary to have a large number available for the emergency of depression conditions. Federal reserve bank notes were originally designed to replace national bank notes. The goal was not achieved in this way, however, and the continued presence of Federal reserve bank notes remains one of the many anomalous elements in the currency system.

There are several currencies issued by the United States Treasury. One of these is the United States note, the correct name for the greenbacks which are still in circulation although in lesser amount than during the War between the States. The Treasury is required by law to maintain a gold reserve of more than one third of the amount of the United States notes, but the Gold Reserve Act of 1934 forbids redemption in gold. Emergency legislation in 1933 authorized the President to issue additional United States notes up to a total of 3 billion dollars. This power was not used, but the greenback continues as an element in the currency system with every promise of remaining in it indefinitely.

Silver certificates are issued by the Treasury, which maintains a silver reserve in its vaults. One-dollar bills are silver certificates, for it is a tradition that bank notes are never in the smallest denominations. Anyone who demands that the Treasury make good its promise to pay the bearer of a silver certificate in silver will be handed a silver dollar in return for a one-dollar bill. A silver dollar is a large and heavy coin, but has it any "intrinsic" value? Actually, the silver in a silver dollar is worth much less than one dollar. The price of silver per ounce in market transactions has for decades been well below a dollar. And a silver dollar weighs less than an

ounce.[9] A silver dollar is worth a dollar not because of the exact amount of silver contained but because it is exchangeable for any other form of dollar; its marks of stamping and milling make it an official currency just as do the engraving and printing on a dollar bill. Far more silver certificates are in circulation than silver dollars. This is a matter of convenience and acceptance. In certain parts of the West silver dollars are used in ordinary circulation, but this is an eccentricity traceable to the traditions of former mining communities.

Besides the silver dollar, there are the other coins, "subsidiary silver coins," and "minor coins" (nickels and pennies). There is little deviousness about the lesser (the dime, the quarter, and the fifty-cent piece) silver coins, for unlike the silver dollar whose weight is still set at 16 times the weight of the pre-1933 gold dollar (a vestige of bimetallism), the subsidiary silver coins have little silver in them. Two fifty-cent pieces or four quarters have less silver than a standard silver dollar. Coins, in fact, of necessity have a higher face value than metallic value, for if the metallic value exceeded the face value the coin would be melted down and sold as metal. Thus silver coins would disappear, as they did in China during the 1930's when the rise in the world price of silver made it profitable to melt coins and export the silver. A government must therefore deliberately maintain the metallic value of its coins at a low level in order to prevent disruption of the coinage system. As a matter of fact, tradition alone dictates the use of silver in the minor coins, since aluminum is actually lighter and cheaper and would thus be more convenient, as several countries have proved. In minting coins, the Treasury makes a "profit," which goes under the traditional name of "seigniorage," signifying the right of a sovereign government to take pieces of metal and transform them into money.

The currency system of the United States is a conglomerate of different things resulting from the experiments, mistakes, and hopes of the past. Only two kinds of currency are really necessary: coins and notes, with one type of each; but the United States has three types of coins and seven types of currency. The salient fact to remember is that the provision of currency is now only a technical function of the money system as a whole. There are no longer any important problems connected with currency. Such problems did exist in the nineteenth century, because then a much larger proportion of money consisted of currency and hence a smaller proportion

[9] A standard silver dollar contains 412.5 grains of silver, $\frac{9}{10}$ fine. The fine silver in a silver dollar is $\frac{100}{129}$ of an ounce. Thus a fine ounce of silver would have to sell for $1.29 if the silver in a silver dollar is to be worth $1. In the middle of 1947 the price of silver in New York was 61 cents per ounce.

of demand deposits. Redemption is of little moment in view of the financial dominance of government; strong governments no longer finance their own activities by printing paper money. The modern money system is capable of automatic expansion and contraction, according to the actual needs for currency.

During World War II the volume of currency in circulation expanded nearly to 400 per cent of its prewar volume. In 1939 the amount was approximately 7.6 billion dollars; in 1946, over 28 billion dollars. Not only was there a rise in the total national income but a rise in the incomes of many persons who do not use banking facilities; furthermore, the constant traveling of people during the war period caused many to hold their money in currency rather than in the form of demand deposits in banks. The increase in currency was possible because the Federal Reserve System so operates as to permit large increases. Most of the wartime increase was in the form of Federal reserve notes. The Treasury was able to augment its supply of coins and of silver certificates, because it had accumulated large stocks of silver before the war; but even if the Treasury had not had silver Congress could have authorized the Treasury to supply more dollar bills.

Contraction of the amount of currency in circulation takes place in this way: when individuals and business firms have more currency than they need they deposit it in banks, which in turn deposit surplus currency in the Federal reserve banks. Since most currency consists of Federal reserve notes, the receipt of their own notes by the Reserve banks causes the Federal reserve notes to cease to exist as money. An I O U in the hands of its maker is nothing—only in the hands of another person is it something. Hence, contraction of currency takes place spontaneously and automatically.

SUMMARY OF THE DEVELOPMENT OF MODERN MONETARY SYSTEMS

The development of money from the earliest use of gold to the system now in effect can be briefly summarized. At first, lumps of gold and gold dust were apparently used as media of exchange as well as measures of value and hoards of value. Under such conditions the total supply of money in a community was equal to the physical amount of gold and could change only if the volume of gold altered. Later, the gold coin appeared, representing nothing more than a certification that a lump of gold was actually of a certain size and purity. Gold coins were much more convenient than gold lumps since they could be accepted at face value, whereas each lump

had to be examined carefully with every transaction. Where coins were the only form of money the supply of money was determined by the number and the sizes of the coins. What frequently happened, however, was that the coinage authority—usually a financially necessitous prince—would call in all coins it could and then reissue them with a smaller gold weight. Such operations were carried out many times in the course of centuries, so that the gold weight of the principal European money units diminished and the supply of money was increased. Hence, where money consists only of coins its supply can be increased by the addition of new gold supplies (as happened after the discovery of the Americas) and by manipulation of the coins themselves.

Following the coin the next step in the evolution of money was paper currency. Paper currency in its beginning seems to have been a mere receipt, representing the deposit of gold with someone owning a strong vault, and was used as circulating money because of convenience. The persons who issued these receipts discovered in due course that they could issue more receipts than they had gold, provided that they could always redeem those receipts presented to them. This discovery—sometimes referred to as the "beginnings of modern banking"—created a paper money only partially backed up by gold. In this context the supply of money depended on the amount of gold and upon the amount of paper money in existence. The amount of paper currency in turn depended upon the reserve system in effect, i.e., upon whether the volume of paper money in existence was backed by perhaps a 20 per cent or a 50 per cent gold reserve. Government control of reserve requirements for paper money did not emerge until the nineteenth century, so that the earlier system of gold-backed paper money involved no control over the total volume of money. Even after the emergence of governmental control there were serious problems of control over currency supply.

Bank deposits as money developed along with paper currency, but logically they constitute an advanced phase. Bank deposits, too, have always had some form or other of reserve. Control over the supply of bank-deposit money is exercised through requirements that banks maintain certain reserves against their deposit liabilities and through control over the volume of these reserves in turn. Typically, however, bank deposits were secured by paper currency which was secured by gold coins which represented definite weights of gold. This raises the problems of how bank deposits act as money, how they increase and decrease, how they are controlled, which will be examined subsequently in detail. The present money system of the United States consists—besides currency—of bank deposits which are se-

cured by other bank deposits (in Federal reserve banks), which are in turn secured by paper currency (gold certificates), which is finally secured by gold that is owned only by the government. Other countries, such as Great Britain and Canada, have similar systems.

CHAPTER 9

The Structure of the American Banking System

In order to prepare for the analysis in the next chapter of the American banking system in relation to the behavior of money and its influence on the level of national output, it is necessary to understand the basic structure of the present banking system and how it evolved from the past. Just as the currency system is explicable only in terms of the events and problems of the past, so the present character of the banking system can most clearly be accounted for by its historical development.

Although there are numerous types of banks in existence, this discussion is concerned chiefly with *commercial banks*, which are the source of the major part of the money used in the economy. It used to be said that the functions of a commercial bank were three: discount, deposit, and issue, i.e., making loans, accepting and maintaining deposits, and issuing bank notes to serve as currency. Commercial banks no longer have the issue function, but retain their discount and deposit functions. Commercial banks lend money, but they are not the only business organizations which do. Pawnbrokers make loans and so do building and loan associations. Nor are commercial banks the only place in which one may open a deposit. The Postal Savings System as well as building and loan associations accept deposits. Commercial banks are therefore actually but one type of financial institution.

TYPES OF FINANCIAL INSTITUTIONS

In order to clarify the differences between commercial banks and other financial institutions, the other types of financial institutions will now be described briefly.

Savings Banks

Although commercial banks accept both savings or time deposits, i.e., deposits which are not legally payable upon demand but only on 30 or 60 days' notice, savings banks receive only time deposits. The cash received through deposits is employed by savings banks to purchase mortgages and long-term bonds; this is one means by which savings are guided into investment channels. The savings banks earn interest on the bonds and mortgages they own; they pay a lower interest to their depositors, thus operating on the difference. Savings banks perform several useful functions. They enable individuals to save their funds when and as they are able, and to receive interest on them. Since the officers of the savings banks perform the task of selecting and buying securities, the income from which is the source of the interest paid to the savings depositor, the individual depositor need not concern himself with the intricacies of mortgages or with the movements of the bond market. The savings banks buy many different types of bonds, and by thus diversifying their investments they minimize risks. The individual depositor's funds are not large enough to permit him to do this directly. The savings bank in effect does it for him. State laws attempt to safeguard the interests of the depositor in a savings bank by prescribing the types of securities which these banks may buy. Because savings deposits are not normally transferable by check, and because they are not payable upon demand, such deposits are not to be considered as money. Therefore the activities of savings banks do not affect the amount of money in existence.

Building and Loan Associations

These associations are similar to savings banks in that they represent a method for the pooling of the financial resources of individuals for mutual benefit. Individuals deposit cash ("purchase share accounts") with these associations, which use the cash to buy mortgages on real estate. As in savings banks, these deposits are individual savings and the purchase of

mortgages is an act of investment. State governments supervise building and loan associations; in addition, the Federal government, through the Federal Home Loan Bank Board, charters and supervises "federal savings and loan associations" which also make real-estate loans and accept deposits from individuals. Because their deposits are also not subject to transfer by check and are likewise not payable on demand, the activities of building and loan associations do not influence the quantity of money.

Finance Companies

These companies make loans for the purchase of automobiles and other durable goods as well as loans to individuals in small amounts. Finance companies are often erroneously referred to as "banks." Actually, they do not have deposits of any kind, since the funds they lend are obtained from the owners of such companies or are borrowed funds, often obtained from commercial banks. A pawnbroker makes loans, but because he does not accept deposits of cash he is not a banker.

Industrial Banks

Banks of the Morris Plan type are now usually included among commercial banks, because most of them do a general banking business including the performance of the special functions of commercial banks. Formerly, however, the industrial banks engaged primarily in maintaining time deposits and making small loans to individuals, loans repayable in installments. Thus, like savings banks, their activities were basically different from those of commercial banks.

Investment Banks

These are businesses which engage in the financing of the new capital requirements of corporations. When a corporation needs money to pay for long-term expansion of its plant and equipment, the corporation may issue new stocks or bonds or both. The stocks or bonds must be sold to investors, either individuals or other businesses. Usually the issuing corporation has no facilities to market its new securities and must arrange with an investment banker to handle the issue. The essential function of the investment banking house, then, is to act as a wholesaler for new issues of stocks and bonds. Investment banks sell new securities to investors, or to other investment banks who in turn sell the securities to investors. Investment

banks do not maintain deposits, although some of them did before 1933, when legislation was adopted definitely segregating commercial and investment banking. Investment banks are of course important to the investment activities of the economy because they introduce, in a sense, those who have savings to those who need money for capital purposes.

Government Lending Agencies

The United States government operates a number of lending agencies, some of which are called banks. Many of these make loans to farmers under the supervision of the Farm Credit Administration. Federal Land Banks make long-term mortgage loans but do not accept deposits. The Federal Loan Agency supervises a number of other lending institutions, the largest being the Reconstruction Finance Corporation, which has lent billions of dollars but also has no depositors. On the other hand, the Postal Savings System of the Post Office Department accepts time deposits but does not in the ordinary sense make loans, for its receipts are invested in government bonds.

It is the *commercial bank*, then, which has the dual function of making loans and of accepting demand deposits (which are transferable by check), besides performing other financial functions as well. Most commercial banks also have time deposits, and invest cash so received in government bonds; many of them make long-term real-estate loans in competition with building and loan associations; they also make small loans to individuals in competition with the finance companies. A large commercial bank is, in fact, a kind of financial department store, with departments or divisions for demand deposits, time deposits, business loans, personal loans, real-estate loans, and for the trust business. They have safe-deposit boxes for rent; they buy and sell securities and foreign exchange for their customers; they may act as travel agents; they may be landlords, if certain real-estate loans prove to be unsound and foreclosure becomes necessary.

THE EVOLUTION OF BANKING IN THE UNITED STATES

To reiterate: commercial banks do two things: they make loans and maintain demand deposits. In eighteenth-century America the function of making loans was performed by merchants, colonial governments, and colonizing companies, although there were no banking institutions in the modern sense. The first modern commercial bank was the Bank of North America, established in Philadelphia in 1782. Other banks were chartered

in the immediately following years by the individual states. In 1791 the first Bank of the United States was founded. Chartered by Congress, its stock was held by the United States as well as by private persons. In addition to commercial bank operations, the Bank of the United States acted as fiscal agent of the government, receiving cash for the government and maintaining a deposit for it, as well as dealing for the Treasury in the bonds of the United States. Like the other banks of the period, it issued notes, for at that time a bank would pay out its own notes to a businessman to whom it had given a loan, instead of setting up a deposit on which he could write checks in the present manner.

The first Bank of the United States was important to the development of American finance because of its relatively large size, because of its intimate relation to the government, and because it could have evolved into a modern central bank. A central bank is one which is fiscal agent to the national government, and which supervises ordinary commercial banks. The first Bank of the United States early exercised supervision over other banks by establishing a system under which all bank notes circulated at par. A businessman early in the nineteenth century was not willing to accept a bank note at its face value unless he was sure he could redeem it at its face value in gold or silver. The Bank of the United States saw to it that all bank notes, including those of its own issue, were in fact readily redeemable, by regular presentation of notes of other banks for redemption as they came into the hands of the Bank of the United States, and thus contributed to the smooth conduct of everyday business transactions. Because of the opposition of the other banks, and because it became embroiled in politics, the charter of the Bank was not renewed after its 20-year term in 1811, and it ceased operation. After the financial exigencies resulting from the War of 1812, Congress in 1816 set up the second Bank of the United States. After some early difficulties, the second Bank of the United States functioned well and like the first Bank assumed a position of leadership among the banks of the country. It too encountered political trouble. After a struggle of several years Andrew Jackson forced its discontinuance in 1836.

After the demise of the second Bank of the United States, banking was regulated only by the laws of the several states. In New England, New York, and Louisiana, the systems of regulation were efficient. Notes issued by banks in these states circulated at par and were readily redeemable in gold, for the banks were required by state laws to maintain reserves behind their note issues. These reserves consisted of bonds, gold, and silver coin. In other states, however, regulation was lax. Banks were established in large num-

bers, many of them by unscrupulous speculators. "Wildcat" banks were founded in remoter regions for the sake of quick profits to be gained from notes issued by roving speculators. Diversity of regulations coupled with the rapid expansion of banking often by inexperienced or fraudulent persons led to the issuance of a conglomerate of bank notes varying greatly in value. Some notes were worthless; others could in fact be redeemed in gold; still others circulated at a discount, at perhaps 75 or 95 cents on the dollar, because they could be redeemed only with difficulty. Every merchant kept a list of the various bank bills in circulation in his area in order to ascertain the value of each note offered to him in payment. Not only did the heterogeneous mass of bank notes of varying values involve much inconvenience, but they also commonly resulted in losses. Holders of bank notes sometimes found themselves with worthless pieces of paper when the issuing bank failed.

The United States government, after itself losing money in the failure of banks in which it had set up deposits, established in 1846 the Independent Treasury System. Under this system the government received and made payments only in specie, i.e., gold and silver coin, maintaining balances in its own vaults. This was a crude method of handling funds, however, one which interfered with the commercial uses of gold and silver money and one which was expensive and awkward for the government. Again in the 1850's the Federal government resumed the practice, with limitations, of using commercial banks as depositories.

The National Banking System

The War between the States provided the opportunity for much-needed reform of the banking system. The Federal government financed the war by means of taxes, by issuing the greenbacks, and by selling bonds. Partly to provide a market for federal bonds and partly for reform purposes, Congress passed the National Bank Act in 1863. The system of national banks established in the following years regulated American banking practice until 1913, and it still exists although largely modified.

The main goal in the establishment of the national bank system was to furnish the country with a safe and uniform bank-note currency. In 1865 Congress passed a law requiring a prohibitive tax on notes issued by state-chartered banks, with the result that the state banks stopped issuing notes and the national bank note came into existence. What was planned was that, besides the creation of new banks, a state bank could give up its state charter and re-establish itself under a charter as a national bank, submitting

to the conditions provided by Congressional law including the regulations and examinations of the Comptroller of the Currency. The new national bank could then issue national bank notes, providing it bought government bonds and deposited them as backing with the Treasury. The notes themselves were engraved by the Treasury and thus had a uniform appearance. Notes of one national bank circulated on the same basis as all the others, and indeed most persons probably paid little attention to the statement on each note giving the name and location of the issuing national bank.

It was expected that the 1865 measure would force all banks to drop their state charters and become national banks, but the state banks did not disappear. In fact, since 1900 there have been many more state than national banks in existence. (A national bank is designated, of course, "national" by its corporate name; any commercial bank without the word "national" in its name is a state-chartered bank.) One reason state banks did not readily convert into national banks was that during the second half of the nineteenth century the volume of demand deposits grew at a faster rate than the volume of currency. In other words, both business and banking habits changed so that in making a loan a bank was more likely to set up a deposit for the borrower than to pay out its notes. Many state banks found that they could get along very well without issuing notes at all. Furthermore, the conditions attaching to the national charters were stricter than those in most states. National banks, for example, could not lend money on real estate at all until 1913, and only after 1935 were they on a par with state banks in this respect. Moreover, state banks often had little supervision by state authorities.

Since the War between the States, then, commercial banking has been controlled under two systems existing side by side: the national banking system and the state banks of 48 separate jurisdictions. Of the approximately 14,500 commercial banks in the United States, 5,000 are national banks, the remainder being state banks.

It is necessary now to examine the main features of the national banking system during the period between 1865 and 1913, when the Federal Reserve System was established. The national bank note, which beginning in 1865 gave the country a safe and uniform bank-note currency, has been described. The operation of the *correspondent* relationship among banks was also an important feature of the national banking system, and indeed still is, although the Federal Reserve System greatly changed some of the results of the correspondent relationship. A correspondent bank is a bank among whose depositors are other banks. To illustrate: bank A is a small bank situated in a small town. To serve its customers bank A finds it convenient

to maintain a deposit in bank B, which is located in a near-by large city. By writing checks on its deposit in bank B, bank A can buy government bonds, or make other large payments. Bank B, in turn, has a correspondent in New York City and maintains a deposit with the New York bank. The New York bank will probably have a correspondent bank in London, with which it keeps a deposit in English pounds sterling. The four banks are thus linked together. Suppose that someone in the small town decides to import British goods. He goes to bank A to arrange to pay the British exporter. Bank A can facilitate the transaction for him, because it has a correspondent arrangement with bank B, which has one with the New York bank, which has one with the London bank to handle this very thing. Banks carry on considerable business with one another which is facilitated by the correspondent relationship, for one bank can easily pay another by sending it a check.

Under the national banking system soon after its establishment the correspondent relationship was such that interbank deposits tended to pile up in New York City banks. Small banks kept their deposits (or balances) with correspondent banks in larger cities, and these banks kept their correspondent balances in banks in New York. Banks outside New York would direct their New York correspondents to use their balances to make loans to finance the purchase of securities on the stock market. Hence, interbank deposits were involved with the fluctuations and the vagaries of the stock market. The banks maintaining deposits for other banks paid interest on these deposits. It was advantageous for small banks to transfer surplus cash to a near-by large city or to New York for deposit and receive interest on it. Since these were demand deposits, the small bank could demand payment in cash at any time. The fact that the large banks paid interest on their correspondents' balances with them meant that they actively sought the deposits of the small banks. The result was that in periods of financial crisis, when all banks were forced to pay cash to worried depositors, the correspondent system made matters worse, because small banks withdrew cash from banks in large cities.

The national banking laws provided for a system of reserves behind the deposits of all national banks. National banks were, and still are, divided into three classes determined by the size of the city in which they are located: central reserve city banks, in New York and Chicago; reserve city banks, in about 60 of the next largest cities; and all remaining banks, "country banks." Before 1913, the law required that central reserve city banks have on hand an amount of gold or other Treasury currency equal to at least 25 per cent of their demand deposits. The reserve city banks also were required to have

a minimum reserve against deposits of 25 per cent; but only half their minimum reserve had to be in cash on hand; the other half could consist of deposits in central reserve city banks, i.e., with correspondent banks. The country banks' minimum reserve was 15 per cent, of which two fifths had to be cash on hand; the remainder could be in the form of deposits in correspondent reserve city banks or central reserve city banks.

Weaknesses of the National Banking System

The national banking system in effect after 1865 was an improvement over the system which preceded it, but it had serious defects. The national bank note possessed the virtues of safety and uniformity, but it failed to meet the needs of the country because it was "inelastic." The total volume of national bank notes could not be expanded to meet seasonal needs, especially at harvest times: sometimes there was an actual shortage of currency. For many years before the adoption of the Federal Reserve System in 1913, agitation for banking reform centered around the problem of how to establish a bank-note currency which would expand and contract according to the demand for currency, i.e., in accordance with changes in the level of national income and with seasonal changes.

Furthermore, the deposit-reserve system for the national banks was faulty. Each national bank kept its own reserves except for that part of its reserve funds which was deposited with correspondent banks, but no individual or organization was responsible for the reserves of the banks as a whole. As has been said, the correspondent system led to a worsening of confidence during times of crisis for it frequently happened that reserve city and country banks made "runs" on central reserve city banks to reassume their deposits, while the latter would try individually to protect their own positions by holding to their own gold or other cash reserves. Furthermore, the Federal government contributed to the instability of the banking system as a result of the operations of the Independent Treasury. At a time when tax collections exceeded expenditures the Treasury would drain the banks of the gold coins and Treasury currency which were a part of their reserves; at a time when federal expenditures exceeded tax collections, the Treasury would pour cash into the banking system. With many circumstances contributing, the banks of the country on several occasions had to shut their doors and business was paralyzed for days or weeks because of the sudden failure of the money system to function.

The major defects of the national banking system can be summarized by stating that the United States, unlike several other major countries during

the nineteenth century, lacked a central bank, or a central banking system. The functions of a central bank are these: (1) to issue notes; (2) to maintain the gold reserves of the country; (3) to maintain the reserves behind the demand deposits of commercial banks; (4) to act as fiscal agent to the national government; (5) to exercise supervision over the entire money system. Central banking functions can be performed without a central bank. Thus under the national banking laws control of the bank-note issue was exercised by the Treasury, although it is a central bank function. In the United States as well as in other countries during the 1930's some of these functions were transferred to the Treasuries, i.e., directly to the administrative government.

After the depression of 1907-08, agitation for banking reform in the United States was intensified with the result that the Federal Reserve Act of 1913 was passed. This act provided for the Federal Reserve System, which was effectively a reform of the existing banking procedure although not a complete reformation of it, for it actually superimposed new controls on the existing banking structure.

THE FEDERAL RESERVE SYSTEM

The basic structure of the Federal Reserve System may be described briefly. The Board of Governors of the Federal Reserve System exercises top control. There are seven governors, each appointed by the President of the United States for a 14-year term at a salary of $15,000. Appointments are by and with the advice and consent of the United States Senate. Under the Federal Reserve Act and its various amendments, the Board of Governors, whose offices are in Washington, D. C., supervises the Federal Reserve System, which consists of the Federal reserve banks. There are 12 Federal reserve banks, located in Boston, New York, Cleveland, Chicago, Philadelphia, Richmond, Atlanta, St. Louis, Dallas, Kansas City, Minneapolis, and San Francisco. Each of these cities is in a Federal reserve district in which there are one or more branch offices of the Federal reserve bank.

The United States has a central banking system rather than a central bank; instead of a single central bank with branches in major cities, there are 12 co-ordinated central banks. The purpose of Congress in establishing the Federal Reserve System in this form was determined somewhat by a traditional fear of concentrated financial authority. Both the first and second Banks of the United States had aroused political hostility, and even a century later Congress harbored a suspicion of a large national bank. Another reason for the multiple banks is the geographical size and the diversity of economic conditions within the United States. Each Federal

reserve bank is able to handle problems peculiar to its own district with an understanding of local conditions and with the facility of being readily accessible. Thus the Federal reserve banks are to some extent permitted to make autonomous decisions.

The Federal reserve banks are not owned by the Federal government but by those commercial banks which are *members* of the system. Member banks, as they are called, consist of all national banks—required by law to be members—and a number of state banks. At the end of 1946, of the 6,900 member banks 5,007 were national banks and 1,893 were state banks. Since there are 9,040 state banks in the United States, only a fifth are members, but these are generally the larger state banks and more than a fifth of the volume of activities of state banks is represented in the Federal Reserve System. The Federal Reserve Act requires a member bank to buy stock in the Federal reserve bank of its district. The stock subscription required is 6 per cent of the member bank's paid-up capital and surplus; [1] one half of the subscription must actually be paid in. On the stock the member banks receive dividends. At the outset of the Reserve System the member banks were required to pay for their stock in gold, so that a large part of the gold held by the commercial banks was in this way transferred to the Federal reserve banks.

Each Federal reserve bank has nine directors. Six of these directors are elected by the member banks (the stockholders), and three are appointed by the Board of Governors of the Federal Reserve System. Three of the six directors elected by the member banks must be professional bankers; the other three must represent other branches of business. The three directors appointed by the Board of Governors are intended to represent the general public interest. One of these three is known as the Federal Reserve Agent, and he also serves as chairman of the board of directors. The Federal Reserve Agent is the direct and full-time representative of the Board of Governors at each Reserve bank. Among his specific duties is the issuance of Federal reserve notes and the holding of the securities behind the Federal reserve notes. The board of directors appoints a president and vice-president who must be approved by the Board of Governors. The president and the other officers are the actual operating heads of the Reserve banks, for the boards of directors meet only to discuss and establish general policy. Control over the Federal Reserve System is thus a divided control between the federal officers in Washington and the men chosen by banks in the different sections of the country.

[1] The paid-up capital and surplus of a bank is an amount which represents what the stockholders originally paid in, together with the profits retained in the bank.

In examining the functions of the Federal reserve banks, it is important to consider them as bankers' banks. With certain unimportant exceptions,[2] the Reserve banks do not deal with individuals or business corporations at all. Their principal relations are with commercial banks and with the Federal government. Deposits with the Federal reserve banks are maintained by member banks, the United States Treasury, and other federal agencies, foreign governments, and foreign central banks. Of these, member bank deposits and Treasury deposits are by far the most important.

Member bank deposits, also called member bank balances, are the reserves of the member banks behind their own deposit liabilities. Minimum reserve requirements for the demand deposits of member banks are 13, 10, and 7 per cent for central reserve city, reserve city, and country banks, respectively. The Banking Act of 1935 authorized the Board of Governors at its discretion to raise legal reserve requirements up to 26, 20, and 14 per cent. This it did in 1936 and 1937. The requirement for central reserve city banks was lowered to 20 per cent during the war years where it now stands. For time deposits, the requirements are lower: formerly 3 per cent, now up to 6 per cent for all member banks. Table 18 gives the legally required reserves.

Table 18 **LEGAL RESERVES**

	Demand deposits		*Time deposits*	
	Maximum	*Minimum*	*Maximum*	*Minimum*
Central reserve city banks	26%	13%	6%	3%
Reserve city banks	20	10	6	3
Country banks	14	7	6	3

If a member bank in a large city has demand deposits of 10 million dollars, i.e., it owes its depositors that much on demand, it must itself maintain a deposit of at least 2 million dollars with the Federal reserve bank in its district. To the member bank its own demand deposits are a liability, while its reserve balance (or deposit) is an asset. To the Federal reserve bank, the 2 million dollars is a liability; it too must be able to pay the amount in cash on demand of the member bank. In practice, member banks keep varying excess reserves on deposit with their Reserve banks, which are amounts more than the legally required minimum. During the latter part of the 1930's total excess reserves were very large.

[2] Under legislation passed in 1934 the Federal reserve banks may make loans to business firms which are unable to borrow from commercial banks but can prove to the Reserve banks that they deserve loans.

A member bank can increase the size of its reserve bank balance in a number of different ways, by depositing currency, by depositing proceeds from the sale of its government securities, or by depositing checks written on one of its correspondent banks. In brief, a member bank can convert some of its assets into money and transfer the money into an increase in its reserve bank balance; similarly, a member bank's balance can be decreased if it withdraws currency from a Reserve bank. The transactions of a commercial bank are such that its reserve balance may be decreased or increased often. The bank must therefore watch its reserve position closely to avoid allowing its reserve balance to fall below the legal minimum.

The reserves of all member banks thus consist of their deposit balances with Federal reserve banks. In addition, a commercial bank always keeps on hand a supply of currency to meet the demands made upon it for cash either by check or deposit withdrawals in currency. The business characteristics of a community (such as the amount and frequency of pay rolls) govern fluctuations in the demand for currency, so that a bank is able to determine easily the amounts to keep on hand. For member banks there are no legal requirements for currency on hand, or till money as it is also called, since legal reserve requirements apply only to the deposits which the member banks must keep with the Federal reserve banks. Thus, the reserves behind the demand deposits of the member banks are centralized in the 12 Reserve banks. Before the Federal Reserve System was established, as already observed, each national bank was responsible for its own reserves and these were therefore scattered in thousands of places. A system of concentrated reserves is more effective than one of dispersed reserves. Percentage reserve requirements have been lower under the Federal Reserve System than they were under the national banking system.

Most advantageously, the Federal reserve banks have so operated that they can provide member banks with additional reserves as needed. One of the basic faults of the national banking system was that the individual commercial bank could not depend on a central bank for aid during a temporary difficulty. Under the Federal Reserve System member banks can borrow from the Reserve banks. Until the 1930's the principal method by which a member bank borrowed from a Reserve bank was by *rediscounting*, which the framers of the original act of 1913 intended, in fact, to be the only method of borrowing. Rediscounting can be illustrated as follows: A wholesaler goes to a commercial bank to borrow money to finance a shipment of goods. The wholesaler wants $100,000 for 90 days, in which time he expects to have sold the merchandise and so repay the loan. To borrow the money from the bank, he signs a note, after providing

adequate assurance that the loan is for the purpose of financing a legitimate commercial transaction. The bank then discounts the note, i.e., it lends the money minus the interest deducted in advance. If the bank charges 6 per cent interest, this would be 1½ per cent for 90 days (a quarter of a year), and on $100,000 the charge is $1,500. Hence the bank lends $98,500, and the wholesaler must repay $100,000 in 90 days. For the sake of this illustration, assume that 30 days later the bank itself needs more money, because the payments the bank has made exceed those it has received and its reserve balance falls too low. (Or the bank could need more money if its own business had grown so much, and its deposit liabilities too, that again its reserve balance is inadequate.) The member bank now goes to the Federal reserve bank and *rediscounts* the wholesaler's note (or "commercial paper"). The note still has 60 days to run, so if the Reserve bank's interest charge ("rediscount rate") is 1½ per cent per annum it deducts ¼ of 1 per cent from the face value of the note. This means that the Reserve bank lends $99,750 to the member bank. This sum is added to the deposit balance of the member bank in the same manner that the member bank, in making the original loan, added the $98,500 to the wholesaler's deposit balance. At the end of the 90-day period, the Reserve bank presents the note to the member bank for payment of the $100,000, and the member bank then presents it to the wholesaler for his payment.

Another way for a member bank to borrow from a Reserve bank is by means of an "advance." This means that the member bank puts up government bonds as a security and borrows on its own note. During the 1930's rediscounting and advances were quite unimportant in the operations of the Federal Reserve System, because the member banks had excess reserves. A member bank must have adequate security to borrow from the Reserve bank: government bonds, or notes signed by businessmen to whom the bank itself has made a loan. In the period from 1930 to 1933 one reason that so many banks failed was that their inability to provide adequate security disqualified requests for aid from the Federal Reserve System. Some commercial banks, however, were given emergency help by the Reconstruction Finance Corporation (established in 1932) which provided funds by buying their preferred stock.

An important technical function of the Federal reserve banks is the clearing and collection of checks. In a small town where there is only one commercial bank there is no large handling problem connected with checks written in the course of the town's business, for most local businessmen will keep deposit balances with the bank and pay one another mostly with checks. The bank simply transfers the funds represented by the checks

from the deposit of one person to another. The use of currency is minimized, and a large volume of transactions is financed by pen and ink on checks and on the deposit ledgers of the bank. In a large city where there are several banks a clearing system is necessary. To examine the simplest case, assume that Mr. A, whose bank is the First National, pays Mr. B with a $100 check. Mr. B's bank is the Second National. At the same time, Mr. C, whose bank is the Second National, pays Mr. D with a $100 check. Mr. D's bank is the First National. Mr. B, the first payee, deposits the check drawn on the First National with the Second National, and Mr. D deposits the check drawn on the Second National with the First National. The two banks owe each other $100. It is obviously simpler for them to reconcile their debts by ledger than to exchange currency. But they do exchange checks, the First National debiting Mr. A's deposit account and crediting Mr. D's, the Second National debiting Mr. C and crediting Mr. B.

In every city where there are several banks, on the other hand, there is a clearinghouse, an organization set up by the banks themselves to facilitate the clearing of checks. The amounts due to and owed by each bank by all the others are compared, and only the difference must be paid by or to a bank. Checks are usually cleared every day. The clearing of checks is another aspect of the convenience of using demand deposits as money, for vast payments can be made by checks drawn on several banks but only small amounts of currency need be used. As a matter of fact, if a bank owes a clearing balance it usually pays it with a check drawn on one of its correspondents or drawn on a Federal reserve bank.

When checks are sent from one city to another the clearing process is more complicated. Assume that a company in New York City receives a check from a company in San Francisco. The check is drawn on a San Francisco bank, but is of course deposited with a New York bank. The New York bank, noting that the check is on a San Francisco bank, deposits it with the Federal reserve bank of New York, which in turn sends the check to the Federal reserve bank of San Francisco for final presentation to the bank on which it was originally drawn. As a result, the reserve balance of the New York commercial bank has been increased, and the reserve balance of the San Francisco commercial bank has been decreased. The New York Reserve bank is owed the amount of the check by the San Francisco Reserve bank. This debt, like a debt between commercial banks, is cleared among the Reserve banks, although instead of through a clearinghouse, final settlement is made through an arrangement called the Interdistrict Settlement Fund.

Finally, while the bulk of intercity checks are cleared through the Fed-

eral reserve banks, some clearing of checks is also done through the correspondent relationships among commercial banks. The efficient clearing system set up by the Federal reserve banks is a tremendous improvement over the cumbersome and slow mechanism which existed under the national banking system. For instance, under the Federal Reserve System member banks and most nonmember banks pay checks at par, but formerly banks were accustomed to making a charge on out-of-town checks, to the annoyance and expense of business.

Besides being bankers' banks, the Federal reserve banks are bankers—"fiscal agents"—to the Federal government. The United States Treasury maintains a large deposit with the Federal reserve banks, as well as with commercial banks. The Treasury's deposit with the Reserve banks acts as a cushion between Treasury financial operations and the commercial banking system, as is shown in the next chapter. Since 1934 the Reserve System has carried on foreign exchange operations and gold transactions for the Treasury. The Reserve banks handle the initial stages of the sale of new Treasury bonds and other marketable federal securities. In addition, the Reserve banks on occasion directly buy the Treasury's short-term obligations, which means in effect that the Reserve banks make short-term loans to the Treasury. Such a practice was formerly regarded as dangerous; for a central bank to lend directly to a Treasury was considered a step toward disastrous inflation.

The functions of the Board of Governors of the Federal Reserve System are concerned mainly with its general supervision of the organization. Within the limits set forth by Congress, the Board can prescribe the reserve requirements of the member banks. It also approves the rediscount, i.e., interest rates charged by Reserve banks to member banks on loans. The seven members of the Board of Governors, together with five committeemen chosen by the Federal reserve banks, constitute the Federal Open-Market Committee. This committee decides for all Federal reserve banks what kinds and amounts of Treasury bonds and other securities the Federal reserve banks shall buy, sell, and hold. Formerly, each Reserve bank was free to decide on its own open-market operations, i.e., the purchase and sale of government bonds by the Reserve banks, but since 1935 these operations have been under the direction of the Federal Open-Market Committee. The committee has also co-operated closely with the Treasury Department in the issue of new securities. Another body which acts with the Board of Governors is the Federal Advisory Council, which is composed of one person from each Federal reserve district. Its purpose is to give the

Board the advice of men experienced in the banking problems of each district, but its importance in the banking system is not great.

Besides its control over the Reserve System and over the member banks, the Board of Governors has still other financial responsibilities. One of these, delegated to the Board by the Securities Exchange Act of 1934, is the regulation of margin requirements in the purchase of securities on the stock market. To buy stocks on margin means that the purchaser puts up money of his own—the margin—and borrows the remainder, the stocks themselves being the security for the loan. So if someone buys stocks with a market value of $10,000 and pays only $5,000 of his own money, borrowing the other $5,000 from a stockbroker or bank, he is operating on a 50 per cent margin. When stocks are bought on margin, the amount bought is larger than the available cash resources of the buyers. Hence margin purchases increase demand, and thus tend to drive the prices of stocks higher in a rising market. The belief that margin buying made the stock market more unstable led to the 1934 measure giving the Board of Governors the power to regulate margin requirements. In recent years margin requirements have been set by the Board at various levels between 50 per cent and 100 per cent. It must be added, however, that the stock market has not gained appreciably in stability through this regulation.

Another regulation that has been enforced by the Board is "Regulation W," governing consumer credit during and immediately after World War II. Under it, the Board prescribed down payments on consumer durable goods, the length of time that loans to individuals could run, the schedule for payment of charge accounts, and similar matters. The purpose during the war period was to limit consumer spending by curbing the amounts consumers could borrow in various ways.

DEPOSIT INSURANCE

During the early years of the depression of the 1930's thousands of banks failed. A bank fails when it is unable to meet its obligations, particularly its obligation to pay its demand depositors in currency. The depositors suffer a loss: their deposits are not convertible into currency and cannot be used for making payments. Usually, the loss to the depositors is not a total one, because the bank has some assets which can be sold to obtain currency to be divided among depositors. The losses suffered by individuals and business concerns during the early 1930's were so great, and such was the general loss of confidence in banks, that Congress was moved in 1933

to set up a system of deposit insurance along with other banking and monetary reforms.

In the history of the banking system there has usually been an effort to protect the public from loss from bad bank notes. Losses did occur before the War between the States, but the national bank note was always safe, whatever its drawbacks from other standpoints. The Federal reserve note is itself surrounded by safeguard after safeguard. The United States government itself has provided directly for the safety of *bank notes.* But not until federal deposit insurance went into effect in 1934 was there direct protection afforded the *bank depositor,* whose money was only in part secured by the requirements of minimum legal reserves. Previously, if a national bank failed the holders of the national bank notes issued by the bank did not lose, for their notes were secured by bonds and by a gold fund maintained by the Treasury; but the depositors lost.

Before 1933 many of the state governments had enacted laws concerning deposit insurance or guaranty, but these efforts to protect depositors were generally not very successful. In the opinion of most observers they could not have been, because the prevention of failures is the proper policy to pursue, rather than the establishment of an insurance fund to pay the depositors of failed banks. But Congress, influenced not so much by state experience as by the spectacle of the losses to depositors, went against conservative opinion and established a system of deposit insurance.

Under the legislation of 1933, as it was amended in 1935, member banks are required and nonmember banks may elect, to insure their deposits with the Federal Deposit Insurance Corporation. The Corporation has three directors, two appointed by the President of the United States, the third one being the Comptroller of the Currency. The original capital of the Corporation was contributed by the Treasury in the amount of 150 million dollars, and by the Federal reserve banks in the amount of 140 million dollars, which represented over half of their accumulated profits. The total capital was approximately 290 million dollars. Each insured bank pays into the Federal Deposit Insurance Corporation an amount equal to $\frac{1}{12}$ of 1 per cent per annum of its total deposits. Deposits of individual depositors are insured by the Corporation only up to $5,000. Hence, if an individual has a $10,000 balance in a bank which fails, he will be paid $5,000 by the FDIC, but the other $5,000 in his case may or may not be paid, depending upon what can be salvaged from the failed bank. Almost every bank in the country, including many not members of the Federal Reserve System, has insured its deposits in this way. The Federal Deposit

Insurance Corporation has the power to examine the banks whose deposits it insures, and it has sought, often successfully, to prevent bank failures. In some cases, it has persuaded a large bank to take over a small and weak one, thus obviating the suspension of the latter. In other cases it has prevented failure by helping a bank over temporary difficulties. Despite this, however, there were 291 bank failures in the period 1935-39 as compared with 8,800 in the period 1930-33; the deposits in these banks were about 126 million dollars, of which 40 million dollars were held by noninsured banks. Thus bank depositors continued to suffer losses, although on a much reduced scale. Bank suspensions dropped off considerably during World War II; in 1945 there were none at all. Since 1934 bank failures can be attributed almost entirely to mismanagement. Of the 14,500 banks in the country, it appears inevitable that every year a few will fail, but these are usually small banks in careless or incompetent hands.

Reasons for Non-Membership

Of the 14,500 banks only about 6,900 are member banks. Why are there more nonmember commercial banks than member banks? There are several reasons for this situation—a situation which limits the effectiveness of the controls exercised by the Federal Reserve System.[3] Many state banks do not join the system because they wish to continue to charge a fee ("exchange") for paying checks sent in to them from out of town; as member banks they would have to pay the checks at face value ("par"). Since the minimum capital requirements of many states are lower than Federal reserve requirements, some small banks are ineligible to become members. Furthermore, the reserve requirements for deposits are not maintained so rigorously in many states as those of the Federal Reserve System, just as in general the regulations applying to member banks are stricter than many applied to state banks. For example, the Federal Reserve Act prohibits a member bank from making a loan of more than $2,500 to an executive officer of the bank. State laws vary on this point but the state requirements are generally easier. Finally, some bankers simply seek to avoid federal regulation on top of the state regulation to which they must submit, although most bankers willingly submit to the regulations concerning federal deposit insurance for the sake of obtaining the advertising value of deposit insurance.

[3] Cf. B. Magruder Wingfield, "Deterrents to Membership in the Reserve System," *Banking Studies*, Board of Governors of the Federal Reserve System, 1941.

BRANCH, CHAIN, AND GROUP BANKING

The banking system of the United States is strikingly different from that of other countries. Perhaps in no single respect is the difference greater than in branch banking. If the American banking system were like the Canadian, allowing for the difference in the size of the two countries, the United States would have possibly 100 banks instead of 14,500. Canada, Great Britain, and France have a few very large banks, with branch offices everywhere. A businessman in these countries deals with a branch of a large bank controlling the affairs of perhaps hundreds of branches.

In the United States, "unit" banking has always dominated. A unit bank is a banking corporation which transacts all of its business in a single office, in contrast to a branch bank which is a corporation which operates several branches in different cities by directing them from a single head office. The advantages of branch banking are: (1) risks are dispersed, so that while a crop failure in one part of the country may cause heavy losses to one branch, it will not affect seriously a whole branch system; (2) a branch bank may develop banking talent and ability to a higher degree, because young men can be given experience progressively through various branches with the result that those who reach the head office tend to be trained, experienced, able, and sound; (3) the managers of branch banks are said to be less likely to make bad loans because they are less influenced by local and personal considerations; (4) branch banks, being parts of a large system, have large resources and can thus meet strains more easily than smaller unit banks. It is true that branch banks and branch banking systems have a commendable record of but few failures. Canada, for example, was as hard hit by the depression of the 1930's as the United States but suffered no bank failures, in contrast to the thousands in the United States, where failures occur even in prosperous times.

On the other hand, proponents of unit banking have claimed that branch banking is inflexible and unadaptable to the special needs of each community; important decisions affecting a locality are made by "absentee" bankers. Under unit banking, local men who know the community thoroughly are the bankers. Above all, unit banking represents a greater degree of freedom of enterprise in banking, for a new bank can be established more easily under this system. It has been objected, however, that the unit banking system enables a few persons who can raise a relatively small capital to start a bank and thus allows many unfit persons to enter the banking business. From the long view it may be recognized that the resentment

in the early nineteenth century against the first and second Banks of the United States reappeared later in opposition to branch banking.

Since the beginning of the 1920's, however, branch banking has grown in many sections of the United States. Over two thirds of the states permit some form of branch banking, whether it is restricted to the head-office city, to the head-office county, or allowed throughout the state. The original National Bank Act did not consider branches at all, but gradually the laws have been changed, so that by the 1930's national banks were in principle permitted to establish branches within states on about the same terms as state banks. The growth of branch banking is indicated by the fact that there were in 1920 approximately 1,500 branches; in 1930, 3,500; in 1946, 4,000 branches. Branch banking is prominent in California and New York. Many of the branch banks are small, operating only one or two branches each.

"Chain banking" means the control of several independently incorporated banks by one or more individuals.[4] Each bank in a chain is a separate legal entity with its own stockholders and directors, but each is controlled by individuals having a controlling stock interest in the chain. Chain banking thus resembles branch banking in that several banking offices are controlled by one unit. In contrast to branch banking, chain banking has invoked only slight attention by state and federal lawmakers. At the end of 1939 there were 96 chain systems in the United States, most of which operated within restricted areas. The greatest concentration of chain banking has been in a few Middle Western states, especially where branch banking is prohibited or restricted. As branch banking has spread, chain banking has tended to decline.

Group banking is still another form of multiple office banking. Several independently incorporated banks are controlled directly or indirectly by a corporation under this system, instead of by individuals, as in chain banking. Usually control is exercised by a holding company which may be the key bank in the system. Like chain banking, group banking has not been subjected to much direct regulation. Under the provisions of the Banking Act of 1933, however, the Board of Governors of the Federal Reserve System has a certain amount of regulatory authority over those holding companies which control member banks. At the end of 1939 there were 41 group banking systems in the United States, chiefly centered in those states which have little or no branch banking.

[4] C. E. Cagle, "Branch, Chain, and Group Banking," *Banking Studies*, Board of Governors of the Federal Reserve System, 1941.

BANKS AND THE CONTROL OF MONEY

The structure of the banking system in the United States is complicated in part because of the country's vast size but mostly because the free enterprise system has allowed banking activities to proliferate and expand while legislative control has been sporadic and diversified. The national banks are subject to a large set of federal laws, the state banks to 48 separate legal codes; the Federal Reserve Act and its amendments applies to the Federal reserve banks, to the national banks, and to some of the state banks. Another aspect of the complexity of law and administration is to be found in the methods of bank examination. A national bank is examined by examiners from the office of the Comptroller of the Currency; it may also be examined by the Federal Reserve authorities and by the Federal Deposit Insurance Corporation. Similarly, a state bank could be examined by two or three different authorities.

It is necessary to emphasize the complexity of banking in the United States, for it is this condition which makes the control of money variable and difficult. It has been shown that the demand deposits of commercial banks are money. The operation of banking therefore determines the supply of money in the economy, but banking decisions and activities are not entirely regulated by the Federal Reserve System, since there are nonmember banks in great number, or entirely by state laws for these are not uniform. The practices of individual banks, of relations between banks throughout the country, of the Federal Reserve System, together with the policies of business, of depositors, of government—these determine the economy's supply of money.

CHAPTER 10

The Functioning of the Banking System

Within the structure of the American banking system, the main banking functions relate to demand deposits. These functions may be analyzed in terms of (1) the relation of reserves to demand deposits, (2) the expansion

and contraction of demand deposits, and (3) the mechanism by which the Federal Reserve authorities control the reserves behind demand deposits. Because demand deposits constitute the larger part of the total supply of money, they constitute the most important element in the working of money systems.

ASSETS AND LIABILITIES OF COMMERCIAL BANKS

To business corporations and to individuals the demand deposits owned by them are assets; to commercial banks, however, their demand deposits are liabilities. Demand deposits are *debts* which the banks *owe* to their depositors. This can be seen in Table 19, a simplified balance sheet of a commercial bank.[1]

Table 19 **STATEMENT OF BANK A**

	Assets	
(1)	Cash (currency on hand)	$ 500,000
(2)	Reserve with Federal reserve bank	4,000,000
(3)	Deposits with other banks	1,000,000
(4)	Government bonds	12,000,000
(5)	Loans	6,000,000
(6)	Bank building, equipment, etc.	200,000
	Total	**23,700,000**
	Liabilities	
(7)	Demand deposits	$18,000,000
(8)	Time deposits	4,000,000
(9)	Capital stock	1,000,000
(10)	Surplus	500,000
(11)	Undivided profits	200,000
	Total	**23,700,000**

First of the assets is cash, or currency on hand, consisting of coins and bills (mostly Federal reserve notes), which the bank requires to cash checks presented to it and to meet withdrawals of deposits. The bank's reserve (2) with the Federal reserve bank is $4,000,000. This reserve is a deposit, or balance, to the credit of the bank. Suppose that the bank is a reserve city bank. Then its minimum legal reserve is 20 per cent of its demand deposits

[1] Assets and liabilities of lesser importance which would normally appear in a bank's financial statement, but which are not relevant to the present explanation, are omitted in this hypothetical statement.

and 6 per cent of its time deposits. Since demand deposits (7) and time deposits (8) are $18,000,000 and $4,000,000, respectively, the bank's minimum reserve requirements must be $3,840,000. Its *excess reserves* are therefore $160,000. The amount of excess reserve in these circumstances is not very large. The bank, however, has deposits with other banks (3) of $1,000,000, and by writing checks on these deposits, can transfer part of them to its Federal reserve balance at any time; if it needs an additional $100,000 in reserve it can write a check on its balance or deposit in the correspondent bank. This reduces the reserve balance of the correspondent at the Federal reserve bank and increases the reserve of the receiving bank. The largest single asset of the bank consists of government bonds (4). Government bonds are sometimes spoken of as the secondary reserves of commercial banks, because government bonds can always be sold by any one bank quickly and with slight risk of serious loss. The item loans (5) is an asset because it represents the promissory notes of businessmen and other borrowers from the bank. The notes, promises to pay specified sums, collectively are usually referred to as the "portfolio" of the bank. The main assets, excepting facilities (6), are also in reality promises to pay: currency on hand, Federal reserve deposit, deposits with other banks, and government bonds.

The assets of the bank, therefore, are what it has—its building, equipment, etc.—and what is owed to it—all other asset items. Its liabilities are what it owes. In the balance sheet, bank A owes $18,000,000 payable on demand to one group of depositors, and $4,000,000 payable after notice to another group of depositors (8). The capital stock item (9) of $1,000,000 represents the original investment of the stockholders of the bank; it is called a liability because the bank, as a corporation, owes this much to its stockholders. Likewise, surplus (10) represents past profits retained by the bank for the ultimate benefit of the stockholders; and undivided profits (11) are those the disposition of which, either as dividends or as surplus, is still to be decided.

The ordinary operations of a bank can now be viewed through their effects upon the balance sheet of a bank. Suppose that a depositor withdraws $100 in cash. In this case both assets (cash) and liabilities (deposits) decline. If the bank's depositors should begin to withdraw currency in large amounts, the bank might find it advisable to add to its stock of currency on hand. It can do this by withdrawing Federal reserve notes from the Federal reserve bank and thereby depleting the reserve balance. If depositors of the bank make large payments by means of checks to business firms which have deposit accounts in other banks, this bank will have

smaller deposit liabilities and also a smaller reserve balance with the Federal reserve bank. The amount by which its reserve balance is reduced is the amount by which other banks' reserve balances are increased. Reserves are lost to a bank whenever clearing balances are adverse.

Loss of reserves through an adverse clearing balance can be illustrated as follows: Suppose that on some one day the depositors of bank A write checks totaling $100,000. Of this amount $25,000 was paid to individuals and corporations also having accounts in bank A; $25,000 is therefore simply transferred on the ledgers of the bank from one set of depositors to another. The other $75,000, however, is paid to persons whose deposit accounts are with different banks. These banks present the checks for payment through the clearinghouse. Now, bank A on the same day presents checks amounting to $25,000 through the clearinghouse for payment from other banks. Therefore, the net amount bank A must pay—its adverse clearing balance—is $50,000. It pays this amount by writing a check on its reserve balance with the Federal reserve bank. As a result, its reserves are lower by $50,000; its deposit liabilities to its customers are lower by the same amount. The $50,000 is paid through the clearinghouse to other banks, and their reserves are thus increased.

Suppose, next, that a businessman obtains a loan from the bank. In the previous chapter it was shown that when a bank discounts the note of a borrower, the bank's deposits are increased by the amount which is lent. In making a loan, then, a bank increases its assets, i.e., its loans and discounts, and adds to its liabilities. The bank's assets are larger because it holds the note of the borrower; the bank's deposit liabilities are higher because the borrower now has his money—in the form of a demand deposit. It would be possible of course for a borrower to obtain currency instead of a demand deposit, in which instance the bank would exchange one asset for another: its currency on hand decreases while "loans" increases. To an extent borrowers do take currency, particularly where very small loans are involved, but usually the borrower gets a new deposit. Thus the bank's loans and deposit liabilities both increase.

It is customary to say that banks "grant credit" to borrowers in making loans to them. The addition to his deposit which a borrower gets is often termed *credit*. By the same definition, the sum total of all demand deposits resulting from business borrowing from commercial banks is termed *bank credit*. But this use of the word "credit" is apt to be confusing, and should be avoided in precise discussion. There is, however, this much to be said for the use of the word "credit." Credit means faith in the validity of a promise to pay. If a businessman's credit is said to be good, it is implied

that those who deal with him have faith in his willingness and ability to meet his financial obligations; poor or bad credit implies unwillingness or inability.

In this sense, banks have often been termed *dealers in credit*. This can be illustrated by a bank loan. A bank loan represents the exchange of the credit of a businessman for the credit of the bank. The borrower and the lender exchange promises to pay: the borrower gives the bank his promise to pay by his note, the bank gives the borrower its promise in the form of a deposit on which he can write checks. In this sense, then, when the borrower writes checks he transfers his claim to the bank's promise to someone else. This process of exchanging credit is one of convenience. The borrower could probably exchange credit with a large number of people with whom he deals in business—and of course promises to pay are common in business transactions—but it is simpler to arrange credit with a commercial bank covering his several needs. The credit, i.e., the promises to pay, of commercial banks are generally acceptable at face value.

In making a loan and in thus adding to its deposits, a commercial bank must face the possibility of losing reserves through adverse clearing balances. In the illustration of bank A's loss of clearing balance, suppose that the $100,000 in checks were written by a business firm which had borrowed $200,000 from bank A and was writing checks on the deposit created by the loan. In one day's transactions at the clearing house, bank A had an (adverse) balance of $25,000 ($100,000 less $75,000 owed to it by other banks). The adverse clearing balance can be met by a check drawn on the bank's balance with the Federal reserve bank, or by some other payment in cash or its equivalent. The point is that a deposit created as the result of a loan, which for convenience can be called a loan-deposit, is obtained by the borrower for active use. Some of the checks written by a borrower, it is true, may go to firms which are depositors of the same bank, in which case there is no loss of reserves. But in making loans the bank must be prepared to lose part of its reserves.

It is important to distinguish between *loan-deposits* and *primary deposits*. The loan-deposit has just been explained. The primary deposit, on the other hand, is the result of the acquisition by the bank of a corresponding amount of cash. Thus, for example, if someone opens a new account, depositing cash (currency or a check drawn on another bank) the bank's cash assets increase as much as its deposit liabilities.

The possibility that a bank will lose reserves when it creates loan-deposits is a factor limiting its power to make new loans. This raises the issue of a bank's reserves in relation to its demand deposits, and hence to

its power to increase loans and deposits. If a bank acquires additional reserves through the clearing process, to what extent can it expand deposits? And if all commercial banks together obtain additional reserves, by how much might total deposits increase?

Reserves and the Expansion of Deposits

The reserves required by law of commercial banks are a means of control over the volume of deposits and therefore over all banking functions. The legal reserve requirements, together with the operations of the Federal reserve banks, limit the powers of commercial banks to add to and to diminish total demand deposits.

Now, it is true that if a commercial bank acquires additional reserves it can increase deposits through the expansion of its loans. It is also true that the reserves of a commercial bank are always less than its deposits, as the statement of condition of any bank (see the statement for the hypothetical bank given previously) will demonstrate. *Yet a single commercial bank can expand its loans and thus its deposits only by amounts about equal to new reserves acquired.*

There is a paradox here: On the one hand, deposits are always several times as large as reserves; on the other hand, a single bank cannot increase loans and deposits by more than the amount of any new reserves it gains. The apparent paradox has been the source of much unnecessary controversy. Bankers have often heatedly denied that banks create money, maintaining that banks lend out "only the dollars entrusted by depositors." Monetary reformers have inclined to exaggerate the ease with which bank deposits are expanded. The paradox can be resolved and the whole question clarified by consideration of two separate cases: (1) a single or isolated bank, and (2) a bank which is one of many in a banking system.

Expansion and Contraction of the Deposits of an Isolated Bank

Imagine a community, isolated from any other, in which there is only one bank. The community is on a gold standard, with gold circulating in the form of coins and with gold as a reserve in the bank behind the paper currency and the demand deposits of the bank. The legal reserve ratio for the bank is 20 per cent, so that for every $100 of deposits and bank notes (currency) outstanding the bank is required to have $20 in gold in its vaults. The money system of the community consists of deposits, currency, and gold coins, deposits being the most important. Hence the people of the

community pay one another largely by means of checks drawn on the one bank; in other words, nearly all payments are accomplished by transfers between deposits in the books of the bank.

The deposits of the single bank in the isolated community can increase if coins or currency are deposited with the bank. But if customers deposit currency, the currency itself goes out of existence since the currency consists of the bank's own notes. Thus deposits and currency are linked together in such a way that if currency in circulation outside the bank increases, deposits contract; and if currency decreases, deposits expand. For the gold coins the relation is different: when gold coins are deposited with the bank, deposits increase but the coins do not, of course, cease to exist—they are added to the reserves of the bank. Besides a reduction in the amounts of coins and paper currency in circulation outside the bank, the important source of deposit increase is the making of loans by the bank. As the bank makes new loans, deposits increase; and as loans are paid off, deposits diminish. An increase in the total deposits of the bank means that the money volume of new loans is greater than the money volume of loans being repaid.

Suppose now that a new gold mine is discovered in the isolated community. The gold recovered from the mine, less whatever amount is used for nonmonetary purposes, such as jewelry, will be added to the reserves of the bank. This will be done either directly, through deposits of gold bars with the bank, or indirectly if the new gold is minted into coins. If all or part of the new gold is coined, the coins themselves could be brought directly to the bank for deposit, adding at the same time to the reserves of the bank; or the coins could be used in business payments. Should this happen, coins in circulation would rise to an amount in excess of business requirements, and the excess would be deposited.

With larger reserves the bank is now able to expand its loans, either by accepting requests it had previously denied or by lowering its interest rates, or both. The expansion of loans involves a corresponding expansion of deposit liabilities. The extent of the increase in loan-deposits is given by the reserve ratio. Since the reserve ratio is 20 per cent, new deposits can be created to the extent of five times the addition to reserves. The ultimate increase of new deposits from a given increase in reserves is therefore in proportion to the reciprocal of the reserve ratio.

Note that the isolated bank *creates* new deposits, and in so doing creates money. It can always do this provided it has the necessary reserves. Because it operates on a system of partial reserves, the bank can create deposits which are not fully backed by reserves. With a reserve ratio of 20

per cent, the isolated bank can create five dollars of new deposits for each one dollar of new reserves. The isolated bank need not fear that it will lose reserves, because there is no other bank to which to lose them. Only if gold coins were hoarded could the bank lose reserves. Such, then, would be the relation of deposits (and loans) to reserves in an isolated bank. But no isolated banks exist; hence, an actual multiple-bank system must be considered.

Expansion of Deposits in a Multiple-unit Banking System

The nearly 15,000 banks in the United States are part of an exceedingly complex system. To show how an addition to bank reserves can increase the total deposits of all banks in the system, and in particular, to show that the proportionate increase in deposits can be as great as the reciprocal of the reserve ratio, requires some simplification in illustration. Of course, expansion of total deposits when reserves increase does happen; in fact, such expansion has always been characteristic of banking systems. The real question is just how the expansion takes place. The process can be illustrated by means of an arithmetical example, with certain simplifications. Assume, first, that all banks have the reserve ratio of 20 per cent. Second, assume that all banks are "loaned up," that is, no bank has any excess reserves at all—their existing reserves are exactly 20 per cent of their existing deposits. Moreover, each bank has to turn down requests for loans, but there are borrowers of good standing who are waiting to borrow funds when the banks are able to lend them. Third, assume that the practice of all banks is to maintain a reserve of 20 per cent behind all new primary deposits, and a reserve of 100 per cent behind all new loan deposits. Fourth, assume that if any bank is able to increase its loans, the entire amount is always taken by one borrower, who promptly writes a check on the entire payment to someone whose own bank is a different one.

Now, suppose that a shipment of $1,000,000 of gold is made to the United States. This is one way in which there can be a net increase in total bank reserves. The shipment of gold is made from London; the shipper consigns the gold to bank A in New York. Why is gold sent from London to New York? It is sufficient for present purposes to say that the shipper in London needs American dollars for trading and is buying them with gold. Bank A, then, sets up a deposit to the credit of the Londoner. The gold itself must be turned over to the United States Treasury, and bank A handles this through the Federal reserve bank of New York. Bank A turns over title to the gold to the Reserve bank and receives an increase in its reserve bal-

ance of $1,000,000. The Federal reserve bank transfers the gold to the Treasury, receiving $1,000,000 in gold certificates.

It may be noted in passing that although the gold may have been taken from the ground in South Africa, only to be buried in the ground at Fort Knox, Kentucky, it has nevertheless performed a monetary function. In coming to the United States it increased both member bank reserves and the reserves of the Federal Reserve System. The gold was used in just as effective a way as if it had actually been turned into coins and the coins put into circulation. For, if new coins are added to those already in circulation and if there is no other change, the new coins will be in excess of those required in circulation. They will then be deposited in banks, which will redeposit the coins with the Federal reserve banks. Hence, bank reserves will increase in this way also. The gold in the possession of the Treasury therefore has done its monetary work, even though it is not in hand-to-hand circulation.

To return to bank A: both its deposits and its reserves have been increased by $1,000,000. The new deposit to the credit of the gold shipper in London is a primary deposit. Since the bank's practice is to maintain a reserve of 20 per cent behind such a deposit, it therefore places $200,000 of the new additional reserves behind its new additional deposit. Bank A then has $800,000 of reserves not required behind any deposits, i.e., $800,000 of excess reserves. Bank A now makes a loan of $800,000 adding to its assets the borrower's note and to its liabilities a new loan-deposit of $800,000. Because of the possibility of large adverse clearing balances, the entire reserve of $800,000 is placed behind the new loan-deposit. The borrower writes a check for the entire amount of his deposit and pays the check to corporation X. The latter deposits the check in bank B, which deposits it with the Federal reserve bank. The Federal reserve bank transfers the amount of the check from the balance of bank A to bank B, and forwards the check to bank A. Bank A then deducts the amount from its borrower's account. The deposits of bank A are now $800,000 less, and its reserve balance is $800,000 smaller. Bank B, however, has acquired additional reserves of the same amount, in the form of an addition of $800,000 to its reserve balance as well as the same addition to its deposit liabilities. The clearing process thus transfers reserves of $800,000 from bank A to bank B.

Now consider bank B with the addition of $800,000 to its reserves. Its deposits too have been increased by the same amount. The new deposit is a primary deposit and against it a reserve of 20 per cent is necessary, or $160,000. This leaves bank B with $640,000 of excess reserves. This bank

can now make a loan of $640,000, and therefore create a loan-deposit of like amount. Suppose that the borrower draws a check on the entire amount and pays it to corporation Y, which does business with bank C. This means that bank B loses reserves to bank C, which gains new reserves. The same process can be repeated for bank C and extended to bank D, and so on. Each bank in the series acquires new reserves, and each increases its deposits by making a new loan, as a result of which reserves are lost to another bank.

This process of deposit expansion is summarized in Table 20.

Table 20 **EXPANSION OF BANK DEPOSITS**

Bank	*New reserves acquired*	*New primary deposits*	*Reserves retained*	*New loan deposits*
A	$1,000,000	$1,000,000	$ 200,000	$800,000
B	800,000	800,000	160,000	640,000
C	640,000	640,000	128,000	512,000
D	512,000	512,000	102,000	409,000
. . . . .				
Total		**5,000,000**	**1,000,000**	

In this table of figures only two of the columns are added—"new primary deposits" and "reserves retained." The other two are not totaled because in both cases the totals would have no meaning, since the individual figures alter as the process of expansion goes on. The total of new primary deposits is given as $5,000,000, and for reserves retained as $1,000,000, because this is the total which is "approached" as a limit (in the mathematical sense) by the successive figures, i.e., if the process is continued until all funds available for transfer are used, this total would result. It is obvious that the total for retained reserves is the same amount as the initial reserve coming into the system.

The difference between the isolated bank and the individual bank which is part of a system is clear. *The isolated bank upon acquiring reserves can expand its loan-deposits by fivefold (with a 20 per cent reserve ratio), whereas the bank which is part of the system expands loan-deposits only dollar for dollar of new reserves.* Yet the expansion in the system as a whole is just as great as with the isolated bank.

It must be pointed out that a bank will not necessarily expand its loan-deposits when its reserves increase. To grant loan-deposits, there must be borrowers who are considered good risks. In the 1930's banks in this country had very large excess reserves, but because of the depression only a

small volume of business borrowing was conducted. At that time bank deposits did not increase in the manner represented here.

As was stated earlier, the multiple expansion of deposits in a system is discussed here with highly simplified figures. A study of banking statistics would indicate that reserves coming into the system as a whole, but initially going to a few banks, are scattered or dispersed among all banks. For example, during the 1930's billions of dollars in gold came to the United States, most of it to banks in New York City. The gold increased the reserves of New York banks. The statistics, however, show that the reserves of New York banks were not relatively larger than those of banks elsewhere in the nation. The newly acquired reserves were quickly lost, through adverse clearing balances, to banks in other cities.

THE FEDERAL RESERVE BANKS

The foregoing discussion was centered on the general principles governing the relations between bank reserves and bank deposits. It is necessary further to examine the behavior of bank reserves in the Federal Reserve System. The reserves of member banks are affected by the operations of the Federal reserve banks, by gold imports, and they are also influenced by the financial actions of the United States Treasury. The demand deposits of member banks constitute more than three fourths of all demand deposits in the banking system. Consequently, whatever affects the reserves (and deposits) *of member banks* influences, for most practical purposes, the entire national volume of demand deposits. The Federal reserve banks themselves are therefore the keys to the banking system as a whole.

The composition of the major assets and liabilities of all twelve of the Federal reserve banks can be seen from Table 21.

The figures in Table 21 are hypothetical, but they approximate the actual figures for 1946. Furthermore, the figures given here include only the major and pertinent assets and liabilities.

The gold certificates (1) held by the Federal reserve banks represent gold in possession of the United States Treasury (gold certificates are, of course, held only by the Reserve banks). Any gold received by a Reserve bank is turned over to the Treasury, which pays in gold certificates. Thus, the heavy gold imports of the late 1930's added materially to the gold certificates of the Reserve banks. Gold certificates are the reserves of the Federal reserve banks. By law, these banks must maintain a reserve in gold certificates of at least 25 per cent behind their note liabilities (the Federal reserve notes outstanding) and their deposit liabilities. From 1913 to 1945,

Table 21 **ASSETS AND LIABILITIES OF THE FEDERAL RESERVE BANKS**

Assets	
(1) Gold certificates	$18,000,000,000
(2) United States Government securities	24,000,000,000
Total	**42,000,000,000**
Liabilities	
(3) Federal reserve notes	$25,000,000,000
Deposits:	
(4) Member banks	16,000,000,000
(5) United States Treasury	1,000,000,000
Total	**42,000,000,000**

the reserve requirements were 40 per cent in gold or gold certificates behind Federal reserve notes and 35 per cent in gold or gold certificates behind deposits. The present 25 per cent requirement for both notes and deposits dates from 1945.

The other important asset is United States Government securities (2). These securities consist of bonds, which are the long-term obligations of the Treasury, and of notes and bills, which are the short-term Treasury securities. The bulk of the government securities held by the Federal reserve banks are of the short-term types. Treasury bills which mature and are paid in cash in 90 days are the most important single type of security owned by the Reserve banks. The Reserve banks buy these securities either directly from the Treasury, a procedure permitted by law since 1942, or they buy the securities in the market. Because most securities are of short term there is a constant turnover of these securities. As individual bills fall due and are paid off by the Treasury, they are replaced by other bills. Thus, the Federal reserve banks buy billions of dollars of securities of the Treasury monthly.

The liabilities of the Federal reserve banks are Federal reserve notes and deposits. The Federal reserve note (3) has already been discussed. These notes come into circulation by being paid out by member banks which obtain them from Federal reserve banks. Since member banks can do this only by drawing down their deposits with the reserve banks, it follows that the volumes of Federal reserve notes and of member bank deposits at the Federal reserve banks are interrelated.

The deposit of the Treasury with the Reserve banks (5) is here set down at 1 billion dollars. The Treasury also maintains deposits with ordinary commercial banks; in recent years these deposits have been much larger

than the deposit balance with the Federal reserve banks. The Treasury's deposit balance with the Federal reserve banks is important, however, because variations in its size affect member bank reserve balances, and can thus affect demand deposits. One of the reasons, in fact, for the creation of the Federal Reserve System was to establish banks which would be both bankers' banks and banks of the Treasury. Thus, both the member banks' deposits (4) and part of the Treasury's deposits (5) are placed with the Reserve banks. As will be seen shortly, an increase in the Treasury's balance generally means a decrease in member bank balances.

Member Bank Reserve Balances and Related Items

The reserve balances of the member banks, i.e., their deposits in the Federal reserve banks, are part of a mechanism. The mechanism may be likened to five wheels geared to one another, so that if one turns the others turn. The five monetary items are:

1. Member bank reserve balances
2. The gold stock
3. "Reserve bank credit"
4. Currency in circulation
5. Treasury cash and deposits

The *gold stock* is the gold owned by the Treasury. An increase in the gold stock has the effect of increasing member bank balances. In the illustration (page 231) of the expansion of bank deposits through augmented reserves, it was shown that a gold import raises bank reserves. Similarly, if a mining company sells gold to the Treasury and is paid by check, this too increases member bank balances. The reason is that the mining company will deposit the check with its bank, which will forward the check to a Federal reserve bank, which will credit (i.e., add to) the reserve balance of the member bank and debit (i.e., subtract from) the Treasury's balance. A decrease in gold stock will diminish member bank balances, if the decrease results from an export of gold handled by a bank. Suppose a private company obtains a license to export gold. It buys the gold from the Treasury, paying for it with a check drawn on a member bank. After the check is cleared, the transaction will have reduced the member bank's reserve balance while increasing the Treasury's balance.

Reserve bank credit is a conventional term for the sum total of the government securities owned by the Federal reserve banks, and of the

discounts and advances made by them to member banks. When the Reserve banks add to their holding of government securities by buying them in the bond market—"open-market purchase"—member bank reserve balances are increased. The reason that member bank reserves are increased is that when a Reserve bank buys government securities directly from a member bank, it pays the member bank by adding the amount of the purchase to the member bank's reserve balance. If, on the other hand, a Reserve bank buys government securities through the bond market from an individual or a corporation, it pays for them by a check drawn on itself. The seller deposits the check with his own bank, which redeposits the check with a Reserve bank, and as a result, the reserve balance of a member bank is increased. Thus it makes no difference who sells securities to a Reserve bank, because this transaction always increases the total of member bank reserve balances. Of course, not every member bank is affected—for some may experience no change at all in their reserve balances; it is the total of reserve balances of all member banks which is affected by open-market purchases. Even where the Treasury sells its securities directly to the Federal reserve banks, the effect is the same. In this case, the Reserve banks pay the Treasury for securities bought by adding to the Treasury's balance. The Treasury spends these funds by writing checks payable to individuals and corporations performing services for the Federal government. When these checks are deposited in commercial banks the latter forward the checks to the Reserve banks, which credit the reserve balances of the member banks concerned.

Discounts and advances are also a part of "reserve bank credit." In the previous chapter it was stated that member banks may borrow from Reserve banks either by rediscounting or by obtaining an advance. By borrowing, the member banks acquire larger reserve balances. Or, looking at this from the standpoint of the Federal reserve banks, the reserves of member banks are increased whenever the Reserve banks make discounts or advances. Hence, member bank reserve balances can be *created* by the Reserve banks, either by purchases of securities or by discounts and advances. The assets—government securities and notes representing discounts and advances—acquired by the Reserve banks in creating reserves for member banks thus become "reserve bank credit."

Currency in circulation and member bank reserve balances are directly related. An increase in currency in circulation lowers member bank reserve balances, while a decrease raises them. The reason is that currency in circulation consists mostly of Federal reserve notes. When currency in circulation increases, as it does regularly every year in December, the public obtains the additional Federal reserve notes by drawing down demand

deposits. The member banks get the Federal reserve notes by drawing down their reserve balances with the Federal reserve banks.

In those periods when tax collections exceed the disbursements of the Treasury, individuals and corporations are paying more into the Treasury than they are receiving from it; hence, there is a net inflow of currency and checks to the Treasury. The currency must be provided by the member banks, unless there happens to be a decrease in currency in circulation at the same time, and they can do so only by depleting their reserve balances. The checks sent in to the Treasury are deposited by it with the Reserve banks, which add the amounts to the Treasury's deposit balance, and subtract the amounts from the reserve balances of the member banks on which the checks are drawn. On receiving the checks, the member banks deduct the amounts from their taxpayer-customers' deposit accounts. Hence, when the Treasury's deposit balance with the Reserve banks increases, member bank reserve balances diminish. On the other hand, when Treasury disbursements exceed tax receipts the reverse holds true: the Treasury's deposit balance falls, and member bank balances increase. When, however, the Treasury sells its securities to the Federal reserve banks, the immediate result is an increase in its deposit balance without any corresponding decline in member bank reserve balances.

Thus the four major factors described here effect changes in member bank reserve balances. Each of these factors has been examined individually. Actually, of course, all four of them act together, some moving in one direction and others in another. Some may be changing in such a way as to raise member bank reserve balances, while others act so as to lower them. Whether or not member bank reserve balances increase or decrease in a given period depends on the net effects of changes in the gold stock, in the volume of securities held by the Reserve banks, in currency in circulation, and in the Treasury's balance.

An illustration of the relation between these factors is afforded by the increase of currency in circulation during World War II. At the end of 1939, currency in circulation amounted to approximately $7,500,000,000; and at the end of 1945, it was about $28,500,000,000. The increase of $21,000,000,000 would have depleted member bank reserve balances had there not been an addition of about $22,500,000,000 to reserve bank credit in the same period. The increase in reserve bank credit was caused by purchases of government securities by the Reserve banks. Furthermore, there was a gain of about $2,500,000,000 in the gold stock from the end of 1939 to the end of 1945.

These developments for the period 1939-45 can be summed up as follows:

Acting to decrease member bank reserve balances:

Increase in currency in circulation	$21,000,000,000

Acting to increase member bank reserve balances:

Increase in reserve bank credit	$22,500,000,000
Increase in gold stock	2,500,000,000

Net change:

Increase of $4,000,000,000 in member bank reserve balances. At the end of 1939 member bank reserve balances were $11,500,000,000; and at the end of 1945 they amounted to $15,600,000,000.[2]

Another illustration of these relationships is given by the increase in the gold stock during the period from the end of 1934 to the end of 1939:

Acting to decrease member bank reserve balances:

Increase in currency in circulation	$2,000,000,000

Acting to increase member bank reserve balances:

Increase in gold stock	$9,300,000,000
Increase in reserve bank credit	100,000,000
Decrease in Treasury cash and deposits with Federal reserve banks	100,000,000
	$9,500,000,000

Net change:

Increase of $7,500,000,000 in member bank balances. At the end of 1934 member bank reserve balances were $4,000,000,000; and at the end of 1939 they were $11,500,000,000.

Member bank reserve balances with the Federal reserve banks are *increased* by any one, or a combination, of the following:

1. *Increase* in the gold stock
2. *Increase* in Reserve bank credit
3. *Decrease* in currency in circulation
4. *Decrease* in Treasury cash and deposit balances

Conversely, member bank reserve balances are decreased by opposite changes in the foregoing factors.

Causes of Changes in Member Bank Reserve Balances

The effects of changes in the factors affecting member bank reserves have been considered. The causes of changes must also be considered. Increases in the gold stock are explained by imports of gold as well as by the transfer

[2] Discrepancies in totals are due to the use of rounded figures.

to the Treasury of newly mined gold. Imports of gold have been far the more important factor. Basically, the reason for the enormous rise in the gold stock after 1934 was that the high price ($35 per fine ounce) established by the United States in 1933-34 attracted gold from all over the world to the United States.[3] Furthermore, the political situation in Europe before World War II caused a flight of capital which took the form of gold shipments to the United States. At this point it is important to realize that the Federal Reserve System has no control over the size of the gold stock; it cannot make additions to or subtractions from it. Not only have Reserve authorities no control over the gold stock, but they may be hindered in the exercise of the controls they do possess.

Currency in circulation changes in accordance with the desires of the public to hold currency. Adjustments in member bank balances are almost automatic when currency increases or decreases. If currency in circulation increases, thus diminishing member bank reserves, the Reserve banks can purchase government securities to raise the balances—as was done on large scale during the war.

Changes in Treasury cash and in the Treasury's deposit balance with the Federal reserve banks are caused by variations in the amount of receipts, expenditures, and debt operations. In the time interval between collection and disbursement, the Treasury may cause member bank reserves to fall. In the 1920's this was of some significance because the excess reserves of member banks were very small and the Reserve banks had to engage in what were then considered to be large-scale open-market purchases in order to maintain member bank reserves at a desired level. During the 1930's, however, member banks generally had large excess reserves mainly because of heavy gold imports, so that the Federal Reserve System did not have to undertake any special action to bolster member bank reserves.

The Federal reserve banks thus act as a kind of cushion between Treasury operations and the member banks. Any adverse effects of Treasury finance on member bank reserves can be offset by the Reserve System.

The principal control, therefore, over member bank reserves is exercised through reserve bank credit. By increasing or decreasing their holdings of government securities and by increasing or decreasing their discounts and advances, the Reserve banks can exert the corresponding effects upon member bank reserves. In so doing, the Reserve banks create or destroy reserves for the member banks and thereby prescribe their power to make loans. But there are limits to this control. The Reserve banks cannot force the member banks to borrow; hence discounts and advances cannot be

[3] See Chap. 11, pp. 256-58.

increased at the will of the Reserve authorities. Holdings of government securities, however, can be varied at the discretion of the Reserve authorities, which, however, is not so great as it may seem. The reason is that purchases and sales of Treasury securities are involved. These securities are bought and sold in an open competitive market, and the prices of the securities are very important. To the Treasury these prices are important because they reflect the success of Treasury borrowing. To the holders of government securities their prices are the values of assets. Any very large volume of selling might depress the prices of government securities, embarrassing the Treasury and causing losses to the owners of government bonds, bills, and notes. It follows that the Reserve banks are not likely to sell billions of dollars of government securities, if they should want to reduce member bank balances.

The Reserve System found itself in a dilemma in the immediate post-war period. (1) If it did not support the government security market (especially for short-term securities) the prices of these securities would fall; (2) if it did support the market, the member banks would acquire additional reserves and be encouraged to expand loans to business or to buy long-term government securities. Because government securities make up the major part of most bank investment portfolios, the Federal Reserve System could not dump its holdings of government securities on the market to reduce member bank reserves if such action drastically reduced security prices. And to support the market was inflationary. To cope with the problem, short-term rates were raised by the Treasury beginning mid-1947 to induce commercial banks to hold their short-term securities. The Board of Governors also recommended to Congress that it be given additional authority to require member banks to hold a "secondary" reserve in the form of short-term government securities. Such a secondary reserve requirement would prevent member banks from selling their government securities at will to acquire reserves, and would restore to the Reserve System control over member bank reserves without sacrificing government security prices.

Changes in Reserve Requirements

The Banking Act of 1935 empowered the Board of Governors of the Federal Reserve System to alter the reserve requirements of the member banks.[4] The effect of increasing the legal reserve requirement is to convert excess reserves into required reserves. Thus, for example, in July, 1936, member bank reserves were $5,900,000,000. Of this total, $2,900,000,000 was

[4] See Chap. 9, p. 214.

excess reserves. In August, 1936, the Board raised reserve requirements by 50 per cent. When this measure had taken full effect, in September, 1936, the situation was: total reserves, $6,300,000,000, and excess reserves, $1,900,000,000. Total reserves were growing (as the gold stock increased), but the Board reduced excess reserves by $1,000,000,000. In so doing it limited the *potential* expansion in demand deposits by reducing the lending powers of the commercial banks.

DEFICIT FINANCING AND THE BANKING SYSTEM

During the war demand deposits about trebled in volume. It has already been noted that the amount of currency quadrupled in the same period. The total volume of currency and demand deposits in 1939 was about $35,000,000,000. By the end of 1946 the supply of money in the United States had grown to nearly $110,000,000,000. The primary cause of this increase in the supply of money was the sale of bonds by the Treasury to the commercial banks.

From the standpoint of their effects on the supply of money, sales of bonds by the Treasury can be divided into two groups: (1) sales to nonbank investors, and (2) sales to commercial banks and to the Federal reserve banks. Nonbank investors include individuals, insurance companies, savings banks, and other business corporations. When these investors buy newly issued Treasury securities, the effect is to transfer demand deposits to the Treasury. But when the commercial or Reserve banks buy new government securities the effect is to *create* new demand deposits. This process should be examined more closely. The person buying a bond for $18.75 pays for it either in currency or by check. The Treasury spends this money, so that someone performing services for the government receives the $18.75. The transaction has had the effect of transferring money from one person to another—from an individual lending to the government to another individual paid by the government. The collection and disbursement of taxes also have the same effect. If a wealthy investor buys half a million dollars' worth of a new issue of government bonds, he pays by check, surrendering half a million dollars of demand deposits in return for the bonds. The Treasury gets the half-million-dollar demand deposit, spends it, and a business corporation supplying materials to the government finally receives the half-million-dollar demand deposit. *The national total of demand deposits is unaltered,* but its distribution is changed.

The sale of newly issued bonds to a commercial bank, however, increases the national total of demand deposits. A commercial bank buys a million

dollars' worth of bonds by means of a check drawn on its reserve balance with a Federal reserve bank. The latter, acting as fiscal agent for the Treasury, delivers the bonds to the commercial bank. The Reserve bank, in handling the commercial bank's check, debits its reserve balance and credits the Treasury's balance. The Treasury spends the million dollars by, say, sending a manufacturing corporation a check for the entire sum. The corporation then deposits the check with its own bank, which forwards it to a Reserve bank which, in turn, credits this commercial bank's reserve and debits the Treasury's deposit balance. The transactions are now complete: the first commercial bank has the bonds, but its reserve balance has fallen; the second commercial bank's reserve balance has risen, but so have its deposit liabilities; the Reserve bank's deposits liabilities are unaltered, as is the Treasury's deposit balance; and the Treasury has paid a bill, but by going more deeply into debt.

The important thing here is that *a new demand deposit exists where none was before,* which is not true of the sale of bonds to private persons and corporations in the bond market. There is nothing in the whole set of transactions to offset or reduce demand deposits. On the other hand, the transactions involving the sale of new bonds to an individual or to a business corporation contain both a *minus* and a *plus.* The *minus* is the reduction of demand deposits when the bonds are bought; the *plus* is the acquisition of demand deposits by someone when the Treasury spends the money. But when commercial banks buy new securities from the Treasury there is no *minus* to cancel out the *plus.* The reason for this is that the commercial bank buys the bonds with part of its assets—its reserve balance. The commercial banks' demand liabilities (deposits) are money; hence the bond purchase does not reduce the volume of money. Spending the proceeds, however, does increase demand deposits.

It is to be noted that the creation of demand deposits by the commercial banks as they buy new Treasury bonds is not a conscious deliberate purpose of individual banks. The creation of money in this way takes place in the system of commercial banks, and is a consequence of the nature of the system rather than of individual design. Individual banks are either the purchasers of bonds or the depositaries of Treasury checks written on the proceeds of bond sales, or they are both. But individual banks do not necessarily connect the two processes.

World War II was therefore financed in part by creating money, just as the War between the States was partly financed by creating greenbacks. The process of creating money by selling bonds to commercial banks was just as real as running the printing presses to turn out paper currency. And

yet there was a difference: the new deposits created during the war were not "forced" into existence in the same manner as the paper currency of a financially exigent government. The banks buying the new issues from the Treasury were not by order compelled to do so, although within the necessities of their operation they had little choice but to employ their idle funds by purchasing government bonds. Their idle funds consisted of their reserve balances, or of funds which could be converted into reserve balances. And, as has been shown, the reserve balances of commercial banks are subject to manipulation by the Reserve authorities, who were always careful enough to provide—mainly by open market purchases of government securities by the Federal reserve banks—a volume of member bank reserve balances large enough so that the banks would absorb new government bond issues.

In the realization that bond sales to banks increased the money supply and intensified the inflationary potential, the Treasury and the Reserve authorities tried during the war to sell as many new bonds as possible to nonbank investors. In particular, as everyone remembers, the sale of the small-denomination, Series E, war bonds was pushed. Likewise, the sale of bonds to business corporations and to insurance companies was intensively encouraged. But during the war the borrowings of the Treasury were always on so vast a scale that the nonbank investors could not buy all of the securities offered by the Treasury. Hence large amounts had to be sold to commercial banks.

Finally, also in the second category of bond sales, the sale of securities by the Treasury directly to the Federal reserve banks has the same effect as to commercial banks; it augments the total of demand deposits. It furthermore increases member bank reserve balances, and thus has a double effect: both demand deposits and member bank reserve balances are increased.

DECREASE OF MONEY SUPPLY THROUGH DEBT RETIREMENT

Just as the money supply is enlarged when the Treasury sells bonds to commercial banks, so too the money supply is reduced when the Treasury retires bonds held by commercial banks.

Suppose that the Treasury uses a billion dollars of surplus tax revenue to reduce the national debt by paying off a maturing billion-dollar issue of bonds owned by commercial banks. When the billion dollars is received as tax revenues paid to the Treasury by checks drawn on individual and corporation demand deposits, the Treasury deposit balance is increased by 1 billion dollars, and member bank reserve balances are reduced by the same amount. Now, the commercial banks turn over the maturing bonds to

the Treasury and receive checks drawn on the Treasury's deposit balance with the Reserve banks. Thus the billion dollars are transferred from the Treasury's balance to member bank balances. At this point the transactions end. The Treasury's balance is the same as before, member bank balances are at their original level, but demand deposits remain lower by 1 billion. There is no offsetting transaction to increase them.[5]

Suppose, instead, that the Treasury uses the 1 billion dollars to redeem securities held by the Federal Reserve System. In this case, the Treasury gives the Federal Reserve a check (or a number of checks) and retires the securities; and Federal Reserve assets are reduced by 1 billion dollars. In effect, 1 billion dollars of public debt, Treasury balance, member bank reserves, and demand deposits are wiped out.

When, on the other hand, the Treasury retires bond issues held by nonbank investors, the volume of demand deposits is unaffected, and therefore the supply of money is not diminished. Suppose that John Smith pays $25 in taxes by means of a check. His deposit account is reduced by that amount. If the Treasury uses the $25 to pay off William Brown's maturing savings bond, Brown's deposit in a bank or else his currency on hand is increased. After the transactions are complete, the total amount of money possessed together by both Smith and Brown is the same, although the one now has less while the other has correspondingly more. The same thing holds true for retirement of larger bonds held by nonbank investors with tax proceeds.

The retirement of bank-held federal bonds reduces the volume of demand deposits only when tax proceeds are the source of funds, i.e., when the money is collected from the public. This is the meaning of the statement that debt retirement is "deflationary," i.e., the supply of money is reduced. If the word "deflation" is used to mean a general and pronounced fall in most prices, paying off the national debt with tax surpluses could hardly be a cause of deflation. The amount by which the money supply could be diminished by retirement of bank-held bonds would not be likely to exceed 1 or 2 per cent of the total supply of money in any one year. Such a reduction in money supply could have, by itself, only negligible effects.

The increase in the supply of money during the war was due, as has just been shown, mainly to the sale of bonds to commercial banks. Only to a small extent was deposit expansion caused through business loans by banks. At the end of 1939 the loans of all commercial banks were about

[5] The only exception would be the case in which banks had been denying loans to potential borrowers for lack of excess reserves, and would begin to make business loans which would offset the reduction in the banks' holding of the public debt.

$17,000,000,000, and the demand deposits of all commercial banks were $32,000,000,000. At the end of 1946, however, loans of commercial banks were $31,000,000,000, while their demand deposits were $92,000,000,000. In a period like the 1920's, loans and deposits expanded and contracted roughly by about the same amounts. But in the war period, deposits were created through the deficit financing of the government.

At the end of the war, business corporations and individuals were in a highly "liquid" condition—having nearly three times as much cash as before the war. This fact is of great significance from the standpoint of control, for it means that the money supply can shrink substantially only through Treasury action. The Reserve System could not, supposing that it desired to, reduce total deposits by reducing member bank reserves, because the only means of doing this would be through open-market sales of Treasury securities, which, as was pointed out, may not be possible on a large scale. The Treasury, however, by retiring public debt held by the Reserve Banks, can reduce significantly member bank reserves.

SUMMARY

This chapter has dealt with the functioning of the banking system as it affects the money supply. An individual commercial bank's demand deposits must be considered in relation to its other major assets and liabilities. Of its assets the most crucial is the bank's reserve behind its own demand deposit liabilities. A bank's reserves are depleted when it has adverse clearing balances, just as they are built up when it has favorable clearing balances. A bank's clearing balances, whether adverse or favorable, reflect its relations with other banks, and depend upon check transactions in the course of business.

When a deposit is made with a bank, its assets and liabilities become larger. Reserves and deposits both increase because the depositor—to acquire the deposit—must transfer cash to the bank, thus either adding to the bank's reserves, or providing the bank with something (e.g., currency) which can be immediately converted into additional reserves. When a bank makes a loan, however, the deposit created as a result of the loan is matched not by additional reserves, but by the borrower's note. Loan-deposits, therefore, must be distinguished from primary deposits.

The total volume of demand deposits rises when there is an increase in loan-deposits. But any one bank cannot extend its lending and deposit-creating activities unless it has adequate reserves.

Banks, however, operate under a system of partial reserves, their deposits

exceeding by far the amounts of their reserves. If there were only one bank, it could expand its loans and deposits by a figure equal to the reciprocal of its reserve ratio. An individual bank in a system of banks cannot do this. The system as a whole, however, operates as does an isolated bank; new reserves coming into the system make possible the expansion of total demand deposits by an amount larger than the new reserves, i.e., in proportion to the reciprocal of the average reserve ratio.

The bulk of the demand deposits in the country are those owed by commercial banks which are members of the Federal Reserve System. The reserves of the member banks consist of the balances (i.e., deposits) maintained by them with the Federal reserve banks. The major assets of the Federal reserve banks themselves are gold certificates, government securities, and discounts and advances to member banks. The latter two are by convention called reserve bank credit. Government securities have for some time accounted for most of the total of reserve bank credit. The chief liabilities of the Federal reserve banks are Federal reserve notes and deposits; the deposits fall into two important groups: member bank reserve balances, and the Treasury's deposit balance.

The importance of the actions of the Federal Reserve System lies in the way in which they affect member bank reserve balances, for demand deposits are limited by the size of member bank reserve balances. Member bank reserve balances are increased by gold imports, by expansion of reserve bank credit through the purchase of government securities by the Federal reserve banks, by a diminution of currency in circulation, and by a drop in the size of the Treasury's deposit balance. Opposite movements in these factors have the effect of decreasing member bank reserve balances.

Gold imports and gold exports, currency in circulation, and Treasury operations cannot be controlled by the Reserve authorities. Hence the manipulations of the Federal Reserve System are primarily restricted to variations in reserve bank credit. Thus, when currency in circulation rose sharply during the war, depleting member bank reserve balances, the latter were kept at the desired level by purchases of government securities.

The Board of Governors of the Federal Reserve System has the authority to alter, within limits, the reserve requirements of member banks. Raising reserve requirements changes excess reserves into required reserves, and has the effect of inhibiting potential expansion of demand deposits.

Treasury borrowing, through the sale of government securities, affects the volume of demand deposits. The sale of new bonds to commercial banks raises the total of demand deposits. On the other hand, the sale of the Treasury of bonds to investors other than commercial banks leaves the total

of demand deposits unaltered. Repayment of the national debt would have the opposite effects: repayment of bonds held by commercial banks will reduce demand deposits, while repayment of bonds held by other investors will have no effect upon the volume of demand deposits.

CHAPTER 11

Monetary Issues and Problems

Of all the different elements in the economic process, money has perhaps received the most attention from both theorists and reformers. The workings of the money system, at once mysterious and fascinating, have above all been a focus of interest for reformers, but economists have devoted themselves at length to the study of money and to consideration of ways of improving the functioning of the money system. Money indeed is typically looked upon as the factor upon which all other economic relations depend. Because it circulates, money has often been likened to the blood stream; just as the blood stream is vital to the human body, so has the circulation of money been considered to sustain all parts of the "body" economic.

MONEY AND PRICES

A dollar has been defined as a unit of claim; it was said that the value of a dollar depends upon how many dollars there are in relation to the number of things for which dollars can be spent.[1] This is the problem of the purchasing power of money.

The purchasing power of money, or, in an equivalent expression, the value of money, is simply the amount of goods and services which can be bought with a unit of money. The purchasing power of a dollar is variable: it was higher in 1939 than in 1947; it was higher in 1913 than in 1920; and it appears to have been larger in 1860 than in 1866. A dollar spent for groceries in 1939 bought more than in 1947, the reason being that prices

[1] See Chap. 8, pp. 186-87.

rose during and after the war. Thus, as prices rise, the value of a dollar declines; as prices fall, the value of a dollar goes up.

The whole group of prices considered together is usually referred to as "the price level." If prices generally rise, it is said that the price level has advanced. Movements in the price level are measured by means of index numbers.[2] The most widely used index number of prices is the index number of wholesale commodity prices published by the United States Bureau of Labor Statistics. The base year for this index is 1926; i.e., the level of prices in that year was arbitrarily taken to be 100. In 1933, the index number of wholesale prices was given as 65.9. From 1933 on, prices rose somewhat and the index advanced to 77.1 in 1939. By 1945, it had gone up to 105.8, and at the end of 1947 it was over 150.

The changes in the value of a dollar can be represented arithmetically by simply inverting index numbers. Thus, for example, if the index number rises from 100 to 200, the purchasing power of a dollar declines from 100 to 50 (i.e., $100/200 \times 100 = 50$).

THE QUANTITY THEORY OF MONEY

The relation between money and prices can be approached first of all through "the quantity theory of money." This theory may be expressed as follows: the value or purchasing power of money varies inversely with the quantity or volume of money in existence, other things being equal. Or alternatively, the level of prices varies directly with the volume of money in existence, other things being equal. This theory is one of the oldest economic principles and has engaged writers on economics for centuries. Actually, it is a special case of a still more general principle, namely, that the value of anything depends upon its quantity, with the necessary proviso that other things remain equal.

It is wearisome to keep on repeating the phrase "other things being equal," but this qualification, which is a necessary corollary to any proposition, is of particular importance to the statement of the quantity theory of money. Why will prices rise if the quantity of money is increased? Prices will rise because a larger amount of money offered for the same volume of goods and services can only be spent if prices do in fact become higher. If prices did not rise, some persons with money would not find the goods they sought. To get the commodities they want these persons offer higher prices. To use an illustrative device of earlier writers on the quantity theory of money, suppose that one morning everybody awoke to find that he had

[2] See Chap. 22 for a fuller discussion of index numbers.

twice as many dollars as before. Everybody, or nearly everybody, would want to buy more than he had been consuming. The resulting increase in the demand for goods and services could not fail to increase prices.

The Equation of Exchange

One way in which the quantity theory of money can be expressed is in the form of a simple equation. This is known as the equation of exchange,[3] which is

$$MV = PT.$$

In this equation, M stands for the total amount of money, V stands for the velocity of circulation of money, P represents the level of prices, and T is the total volume of commodities and services exchanged.

The total amount of money spent in the country during any one year gives a figure which is greater than the figure for the average amount of money in existence during the same year. The explanation of this fact is that most of the money units are spent more than once during a year. A single dollar bill could conceivably be used hundreds of times during a year, as it passes from hand to hand and finances a transaction each time. A business firm may never have more than, say, $10,000 in its checking account which may average $5,000 during a year; but this checking account could be so active (additions and subtractions being made often) that it could easily finance $100,000 worth of business during a year. The *velocity of circulation of money* is the total volume of transactions (measured in terms of money) divided by the average amount of money. For example, if the total volume of transactions is 1,000 billion dollars and if the average amount of money during the year is 100 billion dollars, then the velocity of circulation of money is 10. In other words, the average dollar is spent 10 times.

In the equation of exchange, $MV = PT$, then, MV represents the total amount of money spent. The other side of the equation shows the prices and the amounts of commodities and services exchanged. The sum of the prices of all commodities and services, P, multiplied by their amounts, T, is the sum total of what money is exchanged for, PT. Therefore the equation

$$MV = PT$$

is equal to

Total money spent = total money spent.

[3] Cf. Irving Fisher, *The Purchasing Power of Money*, Macmillan, 1911.

Thus, like any equation, the equation of exchange is an expression of identities.

Another version of the equation of exchange is sometimes used:

$$MV + M'V' = PT.$$

In this version, M stands for gold and currency, and V their velocity of circulation; while M' is demand deposits, and V' their velocity. Here the distinction between the forms of money is made, a distinction which is important where one kind of money depends upon the other. For present purposes, it is well to adhere to the simpler form of $MV = PT$, where M means money in the sense previously defined (currency plus demand deposits).

The equation can then be expressed in this form:

$$P = \frac{MV}{T}$$

(To review the factors, P stands for level of prices, M stands for the amount of money, V for the velocity of circulation of money, and T stands for the total volume of commodities and services.) The equation of exchange now shows more clearly the relation of money and of related factors to the level of prices. Thus, prices, P, can be seen to vary directly with the quantity of money, M, directly with the velocity of circulation of money, V, and inversely with the volume of goods and services, T. *If money and velocity increase faster than goods, prices rise; if money and velocity increase more slowly than goods, prices fall.*

Note that three factors bear upon prices in the equation of exchange, and that each is as important as the others. Therefore, while the quantity theory of money is here being considered, there could just as well be a "velocity of circulation theory" or a "volume of commodities theory" as a quantity theory of money.

Observe also that the equation of exchange says nothing about causal (cause and effect) relations. It does not indicate which are active and which are passive factors. The whole discussion so far has implicitly treated money as the active factor and prices as the passive factor, i.e., an increase in money has been referred to as *the cause* of a rise in prices. But any assertion as to causal relations between money, prices, and related factors must be independent of the equation of exchange itself.

Before examining the question of causal relations more closely, first consider some of the actual relations in the past between money, prices, and goods.

The discovery of the Americas was followed by a great influx of gold and silver into Europe. So great was the increase in the quantity of the two metals that prices rose several fold in the sixteenth century. The time between the inflow of the precious metals and the consequent rise in prices was, it is true, a period of several decades, but there is no doubt that the price changes of the sixteenth century were due to the increase in the supply of money. In the nineteenth century, important gold discoveries were made, in California and Australia in the 1840's and 1850's, and in South Africa, the Yukon, and Alaska in the 1890's. In both periods, the newly mined gold was a substantial addition to monetary gold stocks. Both periods introduced long upward movements of prices. Prices rose almost steadily from 1847 to 1873, and again from 1896 to 1914.

Another type of relation between money and prices is found in periods of inflation. In such periods, the quantity of money and the level of prices rise together, although not of course necessarily in the same proportion. In hyperinflation, such as occurred in Germany after World War I, and in Hungary and Greece after World War II, both money and prices rise to fantastic heights.

Periods of generally falling prices, such as from 1815 to 1848, from 1873 to 1896, and from 1920 to 1933, appear to have been characterized by a rate of growth of production which was faster than the growth in the money supply. In terms of the equation of exchange: T grew faster than M, hence P had to fall. Thus, it is possible to provide a rough kind of verification of the quantity theory of money and of the equation of exchange by reference to historical experience, i.e., to known movements of prices and changes in the money supply in the past. It must be emphasized, however, that the verification is loose and approximate. Only changes in prices over long periods can be easily explained in this way. For short periods, attempts to test the quantity theory turn out to be disappointing. As an example, prices began to fall before the quantity of money began to decline in the depression of 1920-21.

Causes of Monetary Change

The important question is what causes the quantity of money to change. The influences causing the quantity of money to change are the real forces behind major price movements. The quantity of money itself is only the vehicle of change.

In what now follows, these influences affecting the supply of money will

be discussed: (1) changes in the gold stock, (2) the effects of government finance, and (3) the effects of business spending.

In the past it was true that variations in the stock of precious metals were the most important single factors causing the great movements in prices. Changes in the physical volume of gold and silver were due in the main to the accident of the discovery of gold and silver mines. In a sense, then, changes—always increases—in the stock of gold and silver were noneconomic in nature, because at odd moments of history the earth happened to yield new quantities of these metals to be added to money supply.

There was still another factor. The stock of gold of any one country could be increased or decreased through gold imports and exports. The latter in turn depended upon the movements of international trade, but were independent of purely domestic economic developments. When a country's commodity exports exceeded its commodity imports, a "favorable balance of trade" resulted and the country received gold imports; it had an "unfavorable balance of trade," it was forced to export gold.[4] Centuries ago, when these matters were first observed, a crude form of the quantity theory of money was developed. One reason that gold exports were regarded as "unfavorable" was that the accompanying decrease in money supply was thought to lower prices and to have a depressing effect. "Favorable" trade balances were thought to stimulate price advances.

Most of the great problems connected with changing prices and their effects have been caused by the monetary consequences of government finance. The greatest increases in the volume of money have always occurred during or just after periods of war, and these increases have taken place in relatively short periods of time. The connection between government finance and the quantity of money is therefore an important one. It has already been shown how a government, by selling new bonds to commercial banks, can cause an increase in the quantity of money. At this point it is necessary to show the connections among government finance, the volume of money, and prices.

In considering the causes of monetary expansion it is desirable to examine the ways in which demand for money reacts upon effective supply. The demand for money is of course subject to change; generally, over a period of time the demands of business, government, and consumers have been for increased amounts of money. Earlier there was discussed in some detail the *aggregate demand,* or aggregate expenditure, for the national output. This aggregate demand is made effective only when the means of payment are available to business, government, and consumers. It is to be expected,

[4] See Chap. 35.

then, that in any period of rising aggregate demand for commodities and services the supply of money will increase. If currency and deposits do not rise their velocity of circulation will increase. Experience shows, however, that money supply does actually increase in a period of substantial growth of the national output.

This point is revealed in statistics on governmental expenditures in this country during the war period. In 1940 Federal, state, and local governments spent 17 billion dollars for goods and services. By 1944 this had risen to 97 billion dollars with virtually the entire increase attributable to the great expansion of war outlays of the Federal government. For present purposes attention is confined to the 80-billion-dollar increase in 1944 over 1940, spent by government in the purchase of war goods from private producers, in the production of goods in government arsenals and plants, and in payment for military and civilian services. As previously noted, government purchases of goods and services absorbed nearly half of the 210-billion-dollar gross national output in 1944. To make these purchases government had to acquire the necessary funds. Some of these were raised by taxation. This meant a transfer of money income from the taxpayer to government with no increase in total money supply. However, over half of governmental expenditures throughout the war were financed by borrowing: in one year alone, this deficit amounted to 50 billion dollars. Government borrowed partly from individuals and corporations, and partly from the commercial banking system. The former also involved merely a transfer of money from private to public hands. Borrowing from the banks, on the other hand, created money.

When this money was spent for war materials, prices of these materials increased somewhat but governmental price control limited the rise. Once spent the money reappeared as income (wages, salaries, profits) in the hands of wage earners and others. With more money in their possession they demanded more goods and services, but governmental controls limited civilian production and prices. When these controls were removed in 1946, the level of prices was raised because of the increased money supply (currency and deposits), the high level of civilian demand, and the war-induced shortages of civilian goods. The creation of additional money by the banking system in response to governmental demands for money to purchase war goods, therefore, increased the amount of money available to the public at large, which in turn stimulated their demand for goods and services and caused prices to rise.

An increase in the demand of business enterprise for capital goods may have a similar influence on the supply of money. Investment, national

output, and national income are expanded following an improvement in profit expectations. As in the case of government, an increase in the demands of business for commodities and services usually involves a demand for loans (bank deposits). The method of creating deposits has been explained in the discussion on banking. It will readily be seen that here, too, monetary expansion occurs as a result of an increased demand for national output.

Monetary expansion, therefore, will occur when the increased demands by government and business enterprise for money are satisfied by the commercial banking system.

Inflation and Deflation

Inflation means a rapid and general upward movement of prices, just as deflation signifies a rapid and general downward movement of prices. But if inflation meant an equal rise in literally all prices, it would be without much significance to anybody. No one would be worse off, and no one would be better off. With an equal rise in all prices each individual would receive proportionally more for the sale of his commodities or services and would pay proportionally more for what he bought. The problems of inflation arise not so much because prices go up, but rather because they rise unevenly. To say, for example, that a period of inflation is serious because the price level has doubled is to make a tacit assumption that some prices have not risen. Thus, if somebody's cost of living is twice as high as it once was, while his money income is exactly the same, he is only half as well off as he was and inflation is indeed serious for him. As a matter of fact, some prices do not rise at all in a period of inflation and others advance relatively little.

There are always fixed financial obligations in the economy. These include long-term debts, life insurance, pension systems, endowments. The long-term debts include government bonds, corporation bonds, and real-estate mortgages. Pension systems include the social security scheme, as well as private and other governmental pension arrangements. For present purposes, all of these long-term fixed obligations can be thought of as representing a whole series of fixed prices. The interest on bonds and mortgages, the payments to the beneficiaries of insurance policies and of the social security plan, the payments to other retired persons are fixed prices paid for some sort of previous service rendered. These prices, with rare exceptions, do not rise in time of inflation or fall in time of deflation.

Periods of inflation and deflation, therefore, must be considered to include one group of prices which remain stable and other groups of prices which

move unevenly. To think in terms of *the* price level, or *the* purchasing power of money, is perhaps a convenient short cut, but it is likely to be misleading. It is misleading because the very divergences in price movements which constitute the problem of inflation are overlooked.

Inflation and deflation, with the important qualification just noted, mean that the value of the dollar is unstable. Much analysis has been devoted to the problem of how to stabilize the value of money. Periods of deflation, as in 1920-21 and 1929-33, are prolific with plans and schemes for raising the general level of prices, with the avowed purpose of achieving stability. In practice, stability usually means some higher level which existed in the immediate past. The reduction in the weight of the gold dollar in 1933 had just such a purpose.

Reduction of the Weight of the Gold Dollar in 1933

From $20.67 a fine ounce, the price of gold was raised step by step in the summer and fall of 1933 until it reached the present figure of $35.00 per fine ounce. This meant, as was explained earlier, a reduction in the weight of the gold dollar from 25.8 grains, $\frac{9}{10}$ fine, to $15\frac{5}{21}$ grains, $\frac{9}{10}$ fine. The objectives behind the administration's decision to raise the price of gold were to lower the value of the dollar on the foreign exchanges, and to raise prices within the United States. The relation between the gold content of the dollar, and the value of the dollar in terms of pounds sterling, francs, and other foreign currencies is a subject which will be undertaken in Chapter 35. The pertinent issue at this point is the relation between gold and prices within the United States.

The purpose of the reduction in the weight of the gold dollar was, as has just been said, to raise prices. In several public statements President Roosevelt stressed this goal. Now how did the administration expect to lift commodity prices by lowering the weight of the dollar? It is necessary to consider the views of one of the President's economic advisers of 1933. This was Professor George F. Warren, whose views on the subject of gold and prices differed from those of most economists.[5] Professor Warren believed in the validity of the quantity theory of money, but his version of it was unusual. To him, changes in prices were due to variations in the physical quantity of gold as well as in variations in production. If the volume of gold and the volume of production increased at the same rate, prices remained stable. If gold increased faster than production, prices

[5] Professor Warren's views are set forth at length in his book, G. F. Warren and F. A. Pearson, *Gold and Prices*, Wiley, 1935.

would rise; but if production increased faster than gold, then prices would fall. Currency and bank deposits were regarded by Professor Warren as of little importance. Gold coins and gold bars were to him the realities affecting the rise and fall of the price level. His ideas were buttressed by a most formidable array of statistical materials, for he was an indefatigable gatherer of quantitative data. And actually, his statistics did seem to bear him out. In the decades before 1914 there was indeed a rough relation between gold stocks, production, and price indexes.

The explanation given by Professor Warren for the fall in prices after 1929 was "the scarcity of gold." Accordingly, he believed that a reduction in the weight of the gold dollar would raise prices. Now the proposal to alter the price level by changing the weight of the gold dollar was not new. A similar plan, but one more carefully thought out, had been suggested by Professor Irving Fisher, who was the formulator of the most widely known form of the equation of exchange. Fisher's plan was based upon the "compensated dollar," a dollar with a varying weight in gold but with substantially constant purchasing power. Fisher proposed that the monetary authorities watch an index number of prices. If the index should rise, they should increase the weight of the gold dollar and thereby reduce the number of dollars constituted by the gold stock. The adoption of such a measure, Professor Fisher held, would force a contraction in the volume of currency and deposits, and therefore prices would have to decline. On the other hand, a fall in the index of prices would indicate to the monetary authorities the necessity of a reduction in the weight of the gold dollar. This would increase the number of gold dollars and bring about an expansion of currency and deposits; hence prices would rise.

Warren's plan to raise commodity prices by reducing the gold content of the dollar was not founded on any connection between gold and money, and thus between money and prices. He believed that gold directly affected prices (but did not show clearly just how), and the administration at this time held to the same belief. Warren apparently thought that commodity prices would rise proportionately with a reduction in the gold content of the dollar, or an increase in the price of gold would cause a corresponding increase in the prices of all other things. To cause prices to double it was necessary to cut the dollar to one half its former gold content, i.e., to double the price paid for gold by the Treasury.

But what happened? The dollar was finally reduced to 59.06 per cent of its former gold content, and the price of gold was increased 69 per cent to $35.00 an ounce. Prices, however, did not rise accordingly, but in fact rose only slightly. The wholesale index for 1934 was 75, compared to 65 in 1933

(1926 = 100). And even this change could be accounted for by other factors.

Evidently something was wrong with the administration's plan of raising commodity prices by increasing the price paid by the Treasury for gold. One of the many critics of the plan suggested that it was like trying to raise the temperature of a room by recalibrating the thermometer. A criticism of the effort to raise prices by monetary methods in 1933 can be made in this way. For the sake of simplicity assume that changes in the volume of *money* determine the level of prices. The gold operations of 1933 had no direct effect on the volume of money at all. Currency in circulation changed very little after June, 1933. There was a negligible change—an increase—in demand deposits from June to December, 1933, but this was part of a trend, for several years after 1933 both currency and demand deposits showed increases. The administration's purpose in its gold price increase was to raise prices, not in the long run, but immediately by a substantial amount.

The actual effect on the money system of the gold policy can be seen in Table 22.

Table 22 EFFECT OF GOLD POLICY ON THE MONEY SYSTEM

(in millions of dollars)

		Reserve bank credit	*Gold stock*	*Treasury cash and deposits with federal reserve banks*	*Member bank reserve balances*
1934	January	2,656	4,036	397	2,764
	February	2,597	7,138	3,448	2,822
1939	January	2,598	14,599	3,553	9,029

These figures show that the reduction in the weight of the gold dollar increased the gold stock, from one month to the next, and also that it increased "Treasury Cash and Deposits with Federal Reserve Banks" by about the same amount. Thus the Treasury "profit" from revaluation of gold is the principal factor accounting for the increase from January to February 1934. But the Federal reserve banks and the member banks were not affected at all. No asset and no liability of any Reserve bank or member bank was altered in amount by the revaluation of gold—except that the Reserve banks had to exchange their gold for gold certificates, and that the member banks had had to turn in their gold and gold certificates for cash, because gold was legally taken out of circulation.

It was not until much later, months and years after prices were to have

been raised, that the reduction in the weight of the gold dollar had effects upon the money supply. And these effects were not direct, but worked only through member bank reserve balances.

THE SILVER PROBLEM

The United States was on a bimetallic standard until 1873, when the free and unlimited coinage of silver was abandoned. Since that time there has been almost constant agitation to restore silver to its former monetary status, or short of that, to raise the price of silver in relation to gold. This agitation has had some measure of success, with the result that both the monetary system and the Treasury of the United States have been burdened with silver.

The market, or commercial, price of silver fell in the 1870's. Because of the abandonment of the free coinage of silver, there was no longer a fixed legal price at which the mint would buy any quantity which might be offered. Other countries than the United States also dropped bimetallism in this period of time. Furthermore, the 1870's saw the discovery of the Comstock lode and of other rich silver deposits in the west. Thus the supply of silver increased while the demand for it was falling; hence its fall in price.

The general level of commodity prices declined during the 1870's; the beginning of the general decrease in prices is, in fact, usually dated from the year 1873. The fall in commodity prices continued until 1896, also a year of some significance for American monetary policy since it was then that Bryan was defeated for the presidency on his platform for the restoration of silver.

The fall in the price of silver was reason enough for the silver interests themselves to urge the free coinage of silver, so that they could sell the metal to the government at the former (and higher) legal price. The decline in commodity prices from 1873 to 1896 led other interests to ally themselves with the silver promoters. Periods of falling prices are characterized by long and deep depressions, during which heavy pressure is exerted on debtors, for they must pay interest and principal as fixed amounts whereas the prices for their own services and goods decline along with other flexible prices. In addition, friction resulted in the period 1873-96 from the needs of the great expansion in the West on the one hand, and the difficulties in borrowing funds on the other. It will be remembered too that the national bank note, though safe, did not exist in sufficient quantities. Hence there was a widespread belief, with some basis in fact, that

a shortage of money existed. Together with the impact of what seemed to be chronic depression, and the troubles of debtors, this belief produced an agitation for more money. The agitation for more money was joined to and aided the interests of the silver producers.

The pressure to "do something for silver" resulted first in the Bland-Allison Act of 1878, under which the Treasury was required to buy at least 2 but not more than 4 million dollars' worth of silver each month. The silver itself was coined into silver dollars. Then in 1890, the Sherman Silver Purchase Act required the Treasury to buy 4,500,000 ounces of silver per month. The silver bought under this act was paid for with Treasury notes, i.e., paper currency. The latter, however, was made redeemable in either silver or gold. A depression in 1893 brought with it heavy withdrawals of gold from the United States to Europe. Silver was sold to the Treasury, which paid in Treasury notes redeemable in gold. Hence the Treasury's gold stock was being depleted and replaced with silver. So great was the alarm over the rapid substitution of silver for gold that late in 1893 Congress repealed the legislation requiring the purchase of silver.

Thus, finally, with the defeat of Bryan in 1896 the political forces favoring silver were defeated, and the general rise in prices which began in the same year created an economic condition in which the silver problem naturally subsided. The gold standard became firmly established.

During World War I the United States sold about 200,000,000 ounces of silver to Great Britain for use in maintaining the coinage system in India. This might have been an opportunity for getting rid of silver which had lain in the Treasury's vaults since the 1890's, but legislation required that the entire amount be repurchased. The depression of the 1930's revived the demand for the remonetization of silver. The fall in prices after 1929 brought with it the revival of the arguments for increasing the amount of money—chiefly to help the farmer and the debtor. It was argued also that the adoption of silver as standard money by the United States would increase trade with silver-using countries, such as China and Mexico. Actually, the demand for an increase in the supply of money was met by the measures taken in 1933 to lower the weight of the gold dollar, and by the provisions for the emergency increase in the volume of greenbacks. Nevertheless, Congress passed a Silver Purchase Act in 1934. The purpose of this act, in its own words, is "that the proportion of silver to gold in the monetary stocks of the United States should be increased, with the ultimate objective of having and maintaining, one-fourth of the monetary value of such stocks in silver." Hence the Treasury has been required to buy silver in very large amounts, but, because of the great increase in the gold stock

after 1934, without much possibility of attaining the stated objective of the Silver Purchase Act. Moreover, the purchases required by this act had the effect of raising the price of silver considerably. This of course benefited the domestic producers of silver. But in addition the price paid by the Treasury for foreign silver went up markedly, so that the world price of silver was forced up by the buying operations of the United States Treasury. Far from bringing advantages to the silver-using countries, the silver policy of the United States presented them with serious difficulties. China, for example, had been on a silver standard. But with the rise in the price of silver in the 1930's it became profitable to melt coins and to ship the silver bullion to the United States and sell it to the Treasury. As a consequence the Chinese monetary system was disrupted.

It has been said, and with truth, that the silver policy of the United States since 1934 is one of the few economic policies which can be unqualifiedly condemned. For almost any economic policy, however bad, something good can be said—but not for the silver policy. The domestic benefits of that policy have been confined to a very few thousand people, and the effects on international trade and relations have been unfortunate. The Treasury has had to pay for the silver. Only in small part have the costs of the silver been covered by the issue of silver certificates, which are lawfully and solemnly "backed up" by the purchased silver. But in any event, the volume of currency in circulation would have increased steadily during the 1930's, and the Treasury could have issued small-denomination currency without the necessity of buying silver.

OTHER PROPOSALS FOR CURRENCY REFORM

The efforts of reformers to improve the functioning of the money system have not been confined to gold and silver. Other commodities have received their share of attention. Wheat, for example, has been suggested as a better standard than gold. Moreover, a collection of staple commodities has also been proposed as the base for the whole money system as something in which money would be redeemed. Even the erg, a unit of energy, has been advocated as the ultimate monetary standard.

Proposals such as these have been advanced with the laudable purpose of achieving stability of prices and thereby general economic stability. The proposals have been based upon the belief that wheat, or a group of staple commodities—or whatever the commodity may be—will form a more stable base than was provided by the gold standard. It is possibly true that some other commodity could be found the physical production of which would

fluctuate less, from year to year and from decade to decade, than the stock of gold. Actually, however, this consideration is not an important one. In the modern economy it has scarcely any significance at all.

There are obvious reasons for the essential unimportance of the nature of the commodity upon which the money system is based. First of all, one part of the money supply—currency—is controlled in volume and in the extent of its fluctuations by mainly automatic factors, as has been previously noted. All of the important issues center about the behavior of demand deposits, and in the present time currency does not present an economic problem. Now the volume of demand deposits is controlled, if at all, by a system of reserves, in which the actions and policies of the Federal Reserve System play an important role. Changes in the gold stock alter reserves, but the effects of changes in the gold stock can be countered by action of the Federal Reserve System to vary reserves or the required reserve ratio. It follows therefore that the present system of control is capable, if only within certain limits, of handling variations in the gold supply. Consequently, a more stable metallic base than gold is unnecessary, because stability of reserves is achievable through the actions of the Federal Reserve System if it is not forced to serve other ends such as the maintenance of low interest rates to ease the Treasury's problem in handling the national debt. In the second place, proposals for the substitution of something else for gold almost always depend upon a crude interpretation of the quantity theory of money. There is a connection between the volume of money and the level of prices, but it is a varying connection and one which cannot with surety be relied upon. Any reform of the money system which depends on variations in the quantity of money to cause predictable changes in prices is bound to be disappointing. Third, gold is still the universally accepted means of payment in international finance, even though no country now uses it as a domestic medium of exchange. Hence, the abandonment of the use of gold by any one country would seriously affect that country's ability to carry on foreign commerce.

Proposals for Banking Reform

The list of monetary reformers is not confined to those who advocate a standard different from the gold standard, or different from what is presently in use, a "managed inconvertible gold-bullion standard." Some reformers would alter the system of controlling demand deposits. Bills have been introduced into Congress which would, for example, have the effect of ordering the Board of Governers of the Federal Reserve System to vary

the total volume of demand deposits so as to maintain a stable level of prices. The difficulty with such proposals is a twofold one, involving both the ends and the means. Is a stable price level really a desirable end in itself? A good case can be made out for a slowly rising level of prices. For that matter, some economists have thought that periods of rapid technological change should be attended by falling prices. In any case a stable or even an unstable price *level* is a deceptive thing, as was shown earlier in the discussion of inflation and deflation. Even if an index of prices were made stable there might still be profound changes in particular prices, some going up and some going down. It is issues like these which most monetary reformers almost invariably overlook. Even supposing that the end of a stable price level is desirable, what of the means? The question of how to stabilize a price level turns out nearly always to depend on the quantity theory of money. And while the theory is useful for some purposes it is not to be relied upon too much; the important factors are those which cause the quantity of money to change.

A reform proposal which deserves serious attention is the "100 per cent reserve" plan. This plan would require a 100 per cent reserve behind demand deposits to be maintained by commercial banks, instead of the partial reserves (20 per cent or less) which they are now required to maintain. The member banks would acquire the additional reserves presumably by turning in their government bonds for notes bearing no interest (i.e., currency of some sort) which would then be deposited with the Federal reserve banks. The purpose of the plan is to give government an exact control over the total volume of money by depriving the commercial banks of any power, singly or collectively, to create money. The plan would therefore take the deposit creation function away from the commercial banks, just as the note-issue function was previously removed. It would leave commercial banks only with the lending function, and they could lend only their stockholders' money plus their undistributed earnings. The *creation* of deposits as loans are made would be stopped. The details of the 100 per cent reserve plan are rather complex, but it is sufficient to say here that the plan is workable and that a government agency would determine the exact money supply. However, it would be a radical reform of the money system, more radical than any of the reforms of the past and certainly more so in its consequences than the dollar devaluation of 1933. It might be argued that so fundamental a change is scarcely worth the result attained—exact control over the total supply of money—since the demand for money is still free to vary in the course of business fluctuations.

It is pertinent, then, to examine the development of the actual policies

of the Federal Reserve System to see what success with what controls has been achieved.

THE ORIGINS OF FEDERAL RESERVE POLICY

Central banking policy in the period before World War I was primarily a policy of protecting gold reserves. The Bank of England was by far the most important of the central banks; its policies and traditions exerted a strong influence upon the other central banks and were influential on the United States in its framing of the Federal Reserve Act of 1913. The gold possessed by the Bank of England was the reserve behind its notes and its own deposit liabilities. Its depositors included the important commercial banks, which counted their deposit balances with the Bank of England as the reserves behind their demand deposit liabilities. Thus the system was similar to that now in effect in the United States. Since Great Britain was on the gold standard, the Bank of England was obligated to redeem its notes and deposits on demand. In practice, the demand for gold came from those who wanted gold for export. During the course of the nineteenth century the Bank of England learned from experience that by raising its rate of interest—its "discount rate"—it could check the demand for gold for export, so that a rise in the discount rate became the means of protecting the gold reserves of the Bank of England. The reasons that a rise in the discount rate maintained the gold reserves of the Bank of England are quite complex, but they can be summed up as follows: when the Bank of England raised its own interest rates, all other interest rates—those charged by other lenders, including commercial banks—were forced up; with higher interest rates, anyone with liquid funds could earn more, hence the incentive to turn liquid funds into gold for export was reduced. In fact, if the rates of interest in London became high enough, there was a stimulus to move gold to London, turn it into money, and lend the money at the currently favorable rates. On the other hand, when conditions were favorable, when the gold reserve was high, the Bank of England would lower its discount rate, causing other interest rates to fall too. Lower interest rates favored the traders in London, and were regarded as a factor promoting a larger volume of business.

Such was the tradition in the early part of the twentieth century. It was an accepted practice by which a central bank controlled its gold reserve and therefore the total volume of demand deposits by manipulating the discount rate—its own interest rate. So powerful was this tradition that it was adopted by countries where it could have little meaning; it persisted

in England even after the underlying conditions for it had been drastically modified.

When the Federal Reserve System was established in 1913, the intention of the framers of the Federal Reserve Act was to create a central banking system similar in its control to that of the Bank of England. It must be understood that there was no slavish imitation of the British system. In the years before 1913 all banking systems were carefully studied, and features of other banking systems than the British were considered. For example, the Canadian bank-note system was a model, in its elasticity of volume, for the Federal reserve note. The clearing system in Germany was investigated thoroughly to find a means of providing an effective and simple clearing mechanism for bank checks in the United States. But the important issue of control over demand deposits was solved by an adaptation to American conditions of the British banking system of the time.

Thus, in the Federal Reserve System commercial banks were required to maintain their reserves with the Federal reserve banks, just as British banks did with the Bank of England. Likewise, the gold reserves of the banking system were held by the Reserve banks. It was intended therefore that gold would be the reserve behind the Federal reserve banks' note and deposit liabilities, and that the deposit liabilities of the Reserve banks would be the reserve behind the demand deposits owed the public by the member banks. It was intended further that the system would be controlled by the interest rate charged by the Federal reserve banks.

Rediscount Policy

In the early period of the Federal Reserve System, consequently, central bank policy was conceived to be largely a matter of "rediscount policy." The rediscount rate is the rate charged member banks by Reserve banks for rediscounting commercial paper. The theory behind rediscount policy can be summarized as follows: If member banks require additional reserves to expand their own activities of lending and deposit creation, they will have to procure the needed reserves by borrowing from the Reserve banks. Such borrowing needs to be controlled, however, both as to quality and quantity. Control over quality signifies that borrowing of member banks from Federal reserve banks shall be based upon sound business activities only. This in turn means that the loans must be secured by adequate collateral, must be for short terms, and must be for bona fide commercial or agricultural transactions. Control over quantity means variations of the rediscount rate: a high rediscount rate will prevent borrowing from be-

coming too large in volume, and a low discount rate will encourage borrowing when there is an insufficient volume. If the quality and quantity controls are properly exercised, (1) member bank borrowing, therefore (2) member bank reserves, and therefore (3) the volume of demand deposits, will be correctly adjusted to the legitimate needs of business. This in turn would restrain unsound business booms and prevent business panics, such as had occurred periodically in the nineteenth century.

This theory of control by rediscount manipulation contained an implicit assumption of automaticity in the relations between money, prices, and the business cycle. Rediscount policy was not regarded as a means of stabilizing prices or of eliminating the business cycle. Rather, it aimed at the operation of a sound and liquid, a "panic-proof," banking system. Price and business-cycle stabilization would come of themselves.

During World War I the Federal Reserve System was preoccupied with the tasks of war finance and consequently the practicability of its controls could not be tested until 1919. The period of 1919-20 was one of a vigorous postwar boom. Prices rose rapidly and so did the quantity of money. Money increased because of the heavy volume of rediscounting, i.e., extensive borrowing of member banks from the Reserve banks. Rediscounting increased member bank reserve balances, which served as the growing base for expanding demand deposits. The Federal reserve banks saw the dangers in the boom of 1919-20; and they did raise rediscount rates, but not very much, for to have done so would have been too "discouraging to business." Although the Federal Reserve System was not able to prevent the large growth in the money supply with its raise in rediscount rates, when the boom came to an end it was without the collapse of the banking system which had earlier marked transitions from boom to depression. This much may at least be said for the workings of the system at that time.

Because of the strength of the Bank of England's tradition, and because the experience of 1919-20 was interpreted favorably, the rediscount policy of the Federal reserve banks was the focus of banking policy in the 1920's. It is not much of an exaggeration to say that rediscount policy in the 1920's was *the* economic policy of the Federal government. The Reserve System was the only public agency at that time actively engaged in influencing the economy as a whole. Whenever the Reserve banks lowered or raised their rediscount rates, financial writers would debate whether or not it was sound or unsound to "encourage" or "discourage" business at that particular time. Because of large imports of gold to the United States in the early 1920's, the Federal reserve banks could not imitate the Bank of England in using rediscount policy to maintain gold reserves. Reserves were very large and

needed no watching. Hence the new criterion of the "effect on business" of Federal reserve policy. With the advantages of the hindsight possible in the 1940's, the business fluctuations from 1922 to 1929 seem mild. But they were real enough then, and in that period the question was whether these fluctuations could be dampened by the right rediscount policy, coupled with close control of the "eligibility" of the commercial paper which banks could offer the Reserve banks for rediscount.

The test of rediscount policy came in 1928 and 1929. These were the years of soaring stock-market prices, of bank loans on stock collateral, and of complaints of heavy diversion of funds to stock-market activities. The Reserve authorities were powerless to prevent the use of banking funds in the stock market; their only resource in attempting to limit stock-market speculation was to put a check on the total of demand deposits, which involved raising the rediscount rate. This, however, they hesitated to do, except mildly, for a rediscount rate high enough to force banks to curtail loans for stock buying would also have curtailed them for ordinary business purposes. The stock market collapsed in 1929, from its own top-heaviness, and not because of the rediscount policy of the Federal reserve banks. After the crash, rediscount rates were lowered, but with little apparent result. Still later, in the early 1930's, rediscount rates were reduced to almost nominal levels. But if this was to "encourage" business, the result was not apparent.

Open-Market Policy

Thus, rediscount policy collapsed ingloriously in the 1930's. During the 1920's, however, another technique of Federal reserve control had emerged. This was "open-market policy." It was explained in the previous chapter that purchases of government securities in the open market by the Reserve banks increase member bank reserve balances, while sales of government securities in the open market diminish member bank reserve balances.[6] The Reserve banks seem to have stumbled onto this technique in the early 1920's, when they bought government bonds for the sake of earning interest. During the 1920's, open-market purchases and sales were carried on to some small extent, partly to cushion the effect of tax collections on member bank reserve balances. But since rediscount policy dominated, not much attention was given to open-market policy.

The evident failure of rediscount policy in the 1930's, however, led to adoption of a definite open-market policy. Heavy purchases of government

[6] See Chap. 10, pp. 236-41.

securities in 1932 and in 1933 resulted in a situation which has continued to the present time. In the Banking Act of 1935 the importance of open-market operations was recognized by placing control definitely in the hands of the Board of Governors of the Federal Reserve System. Previously it was not clear whether the Board or the individual Reserve banks were vested with control of the volume of open-market purchases and sales.

Open-market operations are a method of control far superior to rediscount rate changes. Member bank reserve balances can be directly and precisely manipulated. A rise in the rediscount rate, on the other hand, might or might not reduce member bank borrowing (and hence their reserve balances); and if it did, the effect might take place indeterminately in one month, or in two, or in three. Open-market policy does have, however, a serious defect. For most practical purposes it can work in only one direction, i.e., by purchasing. Open-market sales, which reduce member bank balances, cannot be made except on a small scale. Large-scale open-market sales would probably cause the prices of government securities to fall, and this would embarrass the Treasury and cause losses to the commercial banks and insurance companies, who are the principal owners of government securities.

Another control possessed by the Board of Governors of the Federal Reserve System is the power to alter the reserve requirements of the member banks. This power was granted by Congress in 1935 and is not a strong one, because the Board may vary the legal reserve ratios only within the limits of 13 per cent to 26 per cent, 10 per cent to 20 per cent, 7 per cent to 14 per cent, respectively, for the central reserve city banks, reserve city banks, and country banks. So far as it does go, however, this power is an effective one, for it can convert excess reserves (of member banks) into required reserves, and vice versa. It was used several times during the late 1930's.

CAN THE FEDERAL RESERVE SYSTEM CONTROL THE SUPPLY OF MONEY?

Despite its instruments of control, the Federal Reserve authorities cannot completely control the total volume of money. In the case of paper currency, volume is determined by the habits of the public and the level of national output in monetary terms. The Reserve authorities of necessity must adjust their operations to the required volume of currency. More significantly the volume of demand deposits—the more important component

of money—is not controllable precisely under the Federal Reserve System as it now exists. These are the reasons:

1. One of the lesser factors is the small membership in the Reserve System. There are more nonmember banks than member banks. True, the member banks account for over two thirds of total demand deposits, but the remaining one third or less of demand deposits is subject to the reserve requirements of the various states, not to the requirements of the Federal Reserve System. Thus, even if by controlling reserves the Federal Reserve authorities could manipulate demand deposits, they could influence only two thirds of all demand deposits.[7] The other demand deposits (of nonmember banks) might change in a contrary direction.
2. Any system of *partial* reserves is a system in which control over the deposits resting on the reserves is necessarily imperfect. An increase in reserves for the system, e.g., from gold imports, may or may not result in a multiple expansion of deposits. The expansion can be as great as the reciprocal of the reserve ratio, but it need not be. There is no telling how great it will be under given circumstances, because there are different reserve requirements for different classes of banks, and because the reserves are always partial and because banks may or may not expand to the limit allowed by their excess reserves.
3. The size of member bank reserve balances cannot be controlled exactly. Heavy gold imports in the 1930's could not be fully offset by the Reserve System. If the Reserve banks had sold government securities in sufficient volume to neutralize the effects of gold imports, the consequences (a decline in the prices of government bonds) would have been disastrous for the Treasury. On the other hand, however, the effects on reserves of movements in currency in circulation and of temporary Treasury surpluses can be offset, if they affect member bank balances in a manner adverse to the wishes of the Reserve authorities.
4. The existence of large excess reserves is proof by itself of the inability of the Reserve System to control the volume of demand deposits. The Reserve authorities may want excess reserves to become required reserves as deposits expand. But businessmen may not be borrowing and the banks may not be creating the new deposits, so that excess reserves would be converted into required reserves. Or, even if the

[7] This statement must be qualified by adding that open-market operations influence the reserves available to nonmember banks in an indirect way.

Reserve authorities do not desire an expansion of deposits, the excess reserves persist as a threat of potential expansion.

5. Treasury policies may upset Federal reserve policy. During the war, as was already explained, Treasury borrowing resulted in a great increase in the money supply. Thus Treasury requirements overrode the then established monetary policy. It is possible that national debt retirement in the future, if it were so carried out as to reduce money supply, could conflict with the purposes of Reserve policy. Reference was made in the previous chapter to the fact that the bulk of the present volume of demand deposits was created as a result of the sale of government bonds to commercial banks during the war. The existence of these deposits frustrates the exercise of Federal reserve control.
6. The Reserve authorities do not control the demand for money. An expanded demand for money comes from businessmen in periods of rising investment expenditures, and from the Federal government in time of war. During the war, the Federal government's demand for money trebled the supply of money. This increase was greater than purely monetary considerations would have dictated, but the Reserve authorities could not interfere with Treasury financing. The demand for money for private investment activity and in the absence of government borrowing could, however, be checked: the Reserve authorities could limit the availability of reserves and so limit deposit expansion. On the other hand, providing large bank reserves and low interest rates cannot stimulate the demand for money in a depression period. A revival of investment requires much more in the economic processes than availability of funds and low interest rates.

For these reasons it must be realized that the Federal Reserve authorities have an imperfect control over the volume of money. If their policies do not conflict with those of the Treasury, the Reserve authorities can create the conditions for an increase or a decrease of the volume of money—but no more than this. Now this conclusion is of considerable importance. It means, for one thing, that as the banking system actually operates, it would not be possible to "stabilize" prices by manipulating the quantity of money, supposing that it were desirable to stabilize prices and that it could be done in this way. Above all, the inability to control money means the subordination of monetary policy to other governmental policies.

From time to time, proposals have been made to strengthen the existing controls possessed by the Board of Governors. The Board itself has asked Congress for authority to alter legal reserve requirements within wider

limits, and to continue the wartime restrictions on consumer borrowing. Such controls, however, work effectively only in checking expansion. They cannot stop contraction.

One of the lessons of the 1930's was the failure of monetary policy—control over the volume of money—to accomplish what had been hoped from it. Prices were actually fairly stable in the 1920's; on the whole it was a prosperous decade. There was a strong tendency to attribute price stability and good business conditions to the very fact that the Federal Reserve System existed and had a rediscount policy. The depression of the 1930's proved that economic stability requires more than monetary policy alone.

The quantity of money does vary. Its fluctuations are roughly parallel with the changes in national output. Such has been the experience up to the present time. Only a radical transformation of the money system, such as the adoption of the 100 per cent reserve plan, would make possible the exercise of effective control over the volume of money and, even then, demand for money would, of course, be beyond the control of the monetary authorities.

PART IV

PRICES AND THE ALLOCATION OF RESOURCES

CHAPTER 12

Prices and Price Functions

The basic concern of this and the following nine chapters is the study of what determines *relative* prices. There are two sets of relative prices, the formation of which are examined separately. First, there are end products, or consumer goods, and second, there are factors: capital goods, labor, and land. Chapters 13 to 16 inclusive deal with the determination of the prices of end products, and Chapters 17 to 21 consider the processes by which factor prices are determined. Both groups of prices affect each other, but to simplify the discussion in Chapters 13 to 16, it will be assumed that factor prices are given.

The commodities and services which constitute the national output are *exchanged* in the processes of production before reaching the stage of final consumption. Each act of physical exchange involves an exchange of money. *Prices are the rates, measured in money, at which commodities and services comprising the national output or income are exchanged.* The national output is in reality measured as a sum of prices. The amounts of commodities in finished form and services rendered directly to persons are multiplied by their prices to measure output in dollars. However, price is employed in this discussion to include not only commodity prices in the ordinary sense, but also wages, interest, and rents. Wage rates are prices of labor, interest rates the prices paid for borrowing money, and rents the prices paid for use of land. Likewise, price will here include rates and fares (such

as those charged by railroads), fees (such as those charged by professional men), and commissions (such as those charged by salesmen), and many other money costs.[1]

Prices are not, however, merely a measuring stick to be used in describing the national income or levels of production and employment. Prices are in themselves a determinant of the composition of national output. Their singular importance in the economy is seen in the fact that, together with technology and natural resources, prices are the principal guide to the millions of decisions made as to what to produce.

The economy of the United States is a *price economy,* in which prices are the principal regulators of economic activity or, more particularly, in which prices are the determinants of the allocation of economic resources. In any modern economy the myriad problems relating to taxes, labor, tariffs, etc., are but special aspects of two major factors: the level of employment of resources and the allocation of resources. It will be seen that prices determine in large measure why resources are used as they are, why certain commodities are produced rather than others, and why commodities are produced in particular ways.

An entire economy may be characterized briefly by the type of regulation applied to its operation. Russia is said to have a planned economy, one which is governed primarily by the decisions of a central governmental authority, in contrast to the economies of the United States and Canada which are regulated chiefly by the individual decisions of the producers and consumers in the entire population.

In the United States and Canada these decisions revolve about prices. The term "price economy" is very often given the same general meaning as "market economy," "competitive economy," "free enterprise system," and "capitalism." Of course each of these terms stresses a particular aspect of economic activity: markets, external regulation, or private ownership of economic resources. Thus while prices themselves cannot be considered the measure of the whole economy, yet the price system may be said to represent the chief aspects of this economy.

The functions of the price system are (1) to distribute consumer goods, (2) to direct production, and (3) to allocate labor and other resources. These

[1] Persons employed by others receive wages or salaries, which are the prices paid for their services. The *self-employed* include professional people and others engaged in service occupations—doctors, lawyers, barbers, tailors, fortunetellers—whose gross incomes consist of the sum of the prices paid for services performed by them. The gross receipts of farmers, and of manufacturing, retailing, and other corporations consist also of the sum of the prices received for commodities and services sold by them. It is, in fact, common practice, when writing about the problems of farm economics, to use farm prices in a sense equivalent to farm income.

functions are integral parts of one process. Consumer goods are distributed by means of the price system among individuals. But the goods have to be produced first, and a favorable price ratio, i.e., prices received exceed prices paid, directs their production. But production entails the use of labor, capital goods, and natural resources, and the prices of these are the incomes of individuals, who are all consumers. Thus, consumers buy goods which are produced with resources the prices of which constitute the incomes of the consumers.

PRICES AND THE ALLOCATION OF CONSUMER GOODS

A function which must be performed in any economy is the distribution or allocation of consumer goods. Once these goods are produced they must be distributed among consumers. And more than that, the flow of production must be so organized as continuously to bring to the final stages of distribution the consumer goods intended for consumption by individuals. Thus the distribution of existing consumer goods and the direction of production toward consumption are two tasks to be accomplished. This is true of the economy of a primitive tribe or of an isolated monastic order or of a socialist society or of an economy like that of the United States. Consumer goods can be distributed in three main ways: (1) according to defined principles prescribed by tradition and community customs; (2) by means of prices under conditions permitting the customers to follow their own preferences in spending their money incomes on commodities; (3) according to a system of rationing or other allocation administered by governmental authority.

The first of these methods prevailed in the simple economies of the distant past, particularly in the manorial economy of the early Middle Ages in Europe. As was noted in detail previously,[2] the annual produce of the land was distributed in kind, part to the lord of the manor and his retainers and part to the cultivators of the land, to each according to his community status. The second method is that of price economies, as in the United States. The third method of allocating goods, by a rationing system, could be made use of in a socialist economy. The Soviet Union has in the past made extensive use of rationing systems. During wartime the United States and other countries have employed rationing to some extent as a supplement to the normal method of price allocation.

[2] See Chap. 2, pp. 29-35.

Distribution of Consumer Goods in a Price Economy

The distribution of consumer goods in a price economy is determined by the free and individual choices of the consumers within the limits of their incomes. Because income differences are very great, some persons are free to select a great many different commodities while others are able to select very few. But it is a system of free consumer choice—each consumer has a money income, all of the commodities available have money prices, and the consumer buys what he wants and is able to pay for.

Out of the great stock of commodities available to him, the consumer buys those kinds in the amounts he prefers and no others. Individual tastes differ widely as to both the kinds and amounts, even in a society where tastes are often said to be standardized. Variations in taste, when exercised on the great quantity of different things produced in the economy, result in many individual patterns of consumption. The *price system* brings about a precise adjustment of commodities to consumer preferences and income or, in other words, to individual patterns of consumption.

It must be added that under a price system each consumer can, within the limits of his income, besides buying what he wants and in the amount he wants, also buy the things he desires when he wants them. Perhaps this factor can best be illustrated by wartime practice in which rationing was imposed on several types of commodities. Holders of ration books found it necessary to observe closely the validity dates of stamps, for if they waited too long to buy, the stamps became invalid. Also, for unrationed goods it was necessary to be vigilant of time. Stocks were sometimes insufficient, and often only those who were on hand to buy at the right times were supplied. It will be readily seen that *time* can be a factor in consumer preferences, for under rationing—which was of course a control superimposed on the price system—the consumer's free choice is limited not only in amount but also in the time and frequency of purchases.

Many commodities are produced seasonally but are consumed the year around, such as foodstuffs like eggs and fruits. Other commodities are consumed seasonally but are produced the year around, such as Christmas cards and swim suits. In all such cases, the adaptation of production and consumption to each other is a fairly complex matter. In part it is accomplished by the storage of a *seasonally produced* good until its market demand is high. Likewise, production of a *seasonally consumed* article may be stabilized by placing it in inventory during periods when consumption is low. Seasonal variations in prices adjust seasonal variations in production and

consumption. Egg prices are low in spring when production is at high seasonal levels, and thus encourage consumption; in the fall, when production is low, egg prices are high and thus check consumption. The prices of oranges, tomatoes, apples, and other food products vary during the year—always in a way to adjust consumption to the fluctuating volumes of production.

The *geographical* distribution of commodities is also governed in part by the price system. To be available to consumers, commodities must be shipped from their places of production to every population center in the land. So unobtrusively does the price system bring about the geographical distribution of commodities that this aspect of its functioning is scarcely realized. Except for unusual items, the consumer takes it for granted that the retail outlets with which he deals will always have in stock the articles he is accustomed to buying. Again, wartime practice may be used for illustration. During World War II certain commodities such as meat and butter were available in some cities, but at the same time either very scarce or not available at all in other cities. The reason was that the maximum prices established by the government early in the war failed to correspond, subsequently, to the actual supply. Consequently the maximum prices in some cities were higher than those in others, so some dealers shipped all or most of their products to the high-price cities. For example, when rationing of most meats was lifted during the summer and fall of 1944, maximum prices on the east and west coasts proved to be slightly less attractive to producers than maximum prices nearer producing centers with the result that for several months meat was very scarce on both coasts but ample in the Midwest and other principal producing areas.

The *types, amounts, and proportions* of goods available are also controlled by prices. A simple example which illustrates this function of the price system is the choice of cuts taken from a side of beef. Various steaks—sirloin, club, round, and porterhouse—as well as roasts, spareribs, plate beef, etc., are to be found. It is possible that these cuts could be distributed by dividing the total number of pounds by the number of consumers and allotting each consumer his share of the pounds of beef. Or the consumers could be grouped in a hierarchy of social rank and then receive portions ranging from the best to the poorest in accordance with their rank. Or, some unkind critic of socialism has suggested, the whole animal could be ground into hamburger and distributed by pounds equally. In a price economy, however, each cut of beef has a different price, and the differences are such that the whole carcass is disposed of during the same time period. No cuts go unsold, none are held over to spoil. Those consumers who want

and can afford the better cuts pay more; the poorer cuts are sold at a low price.

Wartime Price Control

A brief examination of the operation during World War II of the system of price control and rationing is helpful because it will highlight the important aspects of the price system. The prices of consumer goods were controlled and many commodities were rationed by the government. Without government control, prices would have risen sharply. Rapidly rising prices distort distribution, because incomes increase unevenly and those whose incomes become larger obtain more goods at the expense of those whose incomes do not rise.

The control of prices during war was, in principle, an easy task, because the government had only to keep prices from rising. Even in preventing prices from rising, however, the government had its difficulties. Whenever a maximum price is fixed—kept at some level lower than that to which it would rise in the absence of control—the amount consumers seek to buy exceeds the amount offered for sale. If the price were not fixed, it would rise, thus eliminating buyers until the amount the remaining buyers are willing to purchase equals the amounts for sale. When under these circumstances prices are held down by government, the distributive mechanism is thrown into disorder: queues form in front of stores, goods are dealt only to favored customers, and black markets mushroom quietly. For basic commodities public policy demands that rationing be instituted. For example, assume that before the war butter sold at 40 cents a pound and that normally 10 million pounds a year were bought by the consumers in a city. The advent of higher wartime incomes would cause, say, 15 million pounds of butter to be bought. If the price were fixed at 40 cents and only 10 million pounds came into the city, demand would exceed supply. If prices had not been controlled, butter might have gone up to 65 cents a pound, eliminating many low-income buyers and thus making demand equal the supply at this price. If the price were held at 40 cents, 10 million pounds would have to be rationed among the people who would at this price want to buy 15 million pounds.

The wartime rationing system necessitated a very large administrative staff, in Washington, in regional offices, and in every community throughout the land. Besides paid governmental employees, large numbers of volunteer workers were required to make the rationing program function. But the price system requires virtually no governmental administrative supervision

for its operation. There are no policy decisions to be made in Washington, there are no directives to the field offices, there are no local boards, there are no acres of filing cabinets, no voluminous correspondence, no accumulating mass of regulations, no governmental red tape. There are no officials to blame, no policies to criticize. Wartime rationing was no doubt politically necessary, and its consequences preferable to uncontrolled inflation. But where the special atmosphere of a war economy is lacking, the pricing system is superior to rationing. Furthermore, no rationing system can adjust the supplies of available goods so closely to consumers' preferences as can the price system. Under the price system consumers get what they want within the limitations of their incomes. Under a rationing system consumers get what is allowed by governmental allocation. In rationing articles such as sugar and coffee, the government has little choice except to allot each person the same quantity. Although these commodities are consumed almost universally, tastes differ enough so that the ration corresponds only to the preferences of a few.

One very strong indication of the efficiency of the price system was its imitation by the "point system" of wartime rationing. Canned goods, as well as meats and fat, were rationed during the war by assigning to each article a certain point value, by allowing the consumer a quantity of points for each time period, and by permitting the consumer a free choice of the various rationed commodities. (In other countries, clothing was also rationed by means of the point system.) The point system, allowing free consumer choice within the group of rationed goods, differs from the price system in that all consumers have the same income (in points to spend). While the point system of rationing resembles the price system, there is one difference worth noting: Under the point system, consumers must be concerned with whether point values will be raised or lowered, whether the validity of the ration stamps will be extended, whether commodities will be added to or stricken from the ration list. Further, there is the constant problem of keeping in mind two sets of prices—money prices and point value.

Black markets tend to be an accompaniment of fixed maximum prices and of rationing. In wartime, the existence of black markets hampers the economic program of the government. Nevertheless, a black market is one in which the price, however high it may be, equates the amount demanded and the amount supplied; thus the black market is more "normal" than a controlled market. In some liberated countries the black market is reported to have taken care of a far larger volume of transactions than the government-controlled market.

PRICES AND PRODUCTION

The function of prices in allocating any given quantity of consumer goods has been shown. This function is linked to another, directing production. The two functions—allocating the supply of produced consumer goods and directing the flow of production of new consumer and producer goods—are closely related, but it is well to consider them separately because in a socialist economy like that prevailing in Russia, different methods are employed in carrying out these functions. A socialist economy can employ a price system as a means of distributing consumer goods, but it may establish a system of production which is not dependent on prices. In Russia, for example, the amounts, types, and qualities of consumer goods produced are determined by a central planning authority. It is expected that the opinions and values of the planners will not coincide with those of all the consumers. Socialist planners might decide to produce more men's shirts and fewer men's socks at a time when additional socks rather than shirts are wanted by the consumers.

In a capitalist economy, on the other hand, production is not directed by central authority but by prices. That is, prices producers expect to get indicate prospective profitability and in this manner direct production. Even in the face of wartime controls industry was sufficiently flexible to respond to changes in consumer demand. For example, when there was a fear of enemy air raids on the coastal areas of the United States, defense measures were undertaken by the government, and very quickly blackout curtains, sand containers, and other equipment appeared in shops to be bought and installed by worried citizens. These things appeared so quickly not because of governmental orders, but because manufacturers sensed a profit and capitalized on it.

The price system, under which producers govern themselves according to consumer demand, has been likened to political democracy in this way: politically the individual votes his approval or disapproval of public policies by casting a ballot; economically, as it were, by spending his dollars. By spending them in one direction he indicates approval, and by refraining from spending them in another direction he signalizes his disapproval. This parallel cannot be carried very far, because while each person has one vote at the polls, the number of dollars at the disposal of individuals varies widely.

In directing the flow of production prices determine (1) the types of goods produced, (2) the quantities produced, and (3) the methods by which goods

are produced. Different types of goods and services are produced according to profit expectations which in turn depend upon price expectations. Every investment decision hinges upon profit expectations which are expectations of future price relations. If manufacturers expect the price of a product to rise, without a corresponding rise in the prices of labor and material, they will produce more of the product.[3] Rising prices for particular commodities result from increases in demand on the part of consumers, either because of a change in their tastes or because their incomes increase. The sequence of increased demand, higher prices, larger profits, and expanded production does not always occur chronologically. Producers may anticipate a change in consumer demand, thus anticipating price and profit changes, and therefore increased *quantities* may arrive on the retail market just as the increased demand is manifested. On the other hand, falling prices are signals to producers to curtail output. Most producers actually produce more than one good, so that a falling price for one good usually results in diverting resources from it to the production of others.

Methods of production are also determined by prices, within the possibilities presented by the state of technology. A farmer uses land, hired labor, machinery, and fertilizer to produce his crops, all of which may be combined in different ways. He can produce a crop of some given size on a large piece of land, utilizing no machinery or fertilizer; or he can combine a small piece of land with intensive application of fertilizer and machinery to produce a crop of the same size. To a considerable extent,[4] a farmer can substitute labor for machinery and machinery for labor. The amounts in which these four resources are employed are dependent on their prices. If land is cheap and labor is dear, more land and less labor will be utilized; if land is dear and machinery is cheap, less land will be used but it will be farmed intensively; etc. In a price economy the price of a resource, such as land or farm labor, is an index of its over-all scarcity. It is desirable for an individual as well as for the nation that cheap resources be used in larger quantities than dear resources. In manufacturing, where there is an even greater range of possible methods of production, the need is likewise to choose the lowest priced combination of labor, materials, and machinery.

Alternative Controls of Production

In the modern world the chief alternative to the control of production by prices is government authority. Even in a price economy the government

[3] See Chap. 5, pp. 111-18, on the role of prices, profits, and expectations.

[4] Subject to the facts of diminishing returns and the indivisibility of the factors of production (discussed in Chap. 17).

already plays an important part in directing production, partly supplementing the results of the price system and partly opposing its free operation. Government prohibits the production of certain *types* of goods and services, such as harmful drugs, some forms of gambling, and sometimes beverage alcohol for nonmedicinal uses. On the other hand, the government produces some services itself, such as public health services, highways, public education, which can be produced under a price impetus though in different form. Government also modifies the price system in the *quantity* of production. Subsidies, for example, are paid to certain producers to increase output.

The government influences methods of production very considerably by certain controls on prices. By establishing minimum rates of wages, and by encouraging collective bargaining which in turn tends to raise wages, contemporary government has impelled the substitution of machinery for labor in many fields of production. By regulating the rates charged by railroads more stringently than the rates of motor carriers and airlines, the Federal government has probably diverted some traffic from the railroads to bus lines, trucking companies, and airlines. By raising the prices of imported commodities through the imposition of tariffs, the government has encouraged the production of some commodities which otherwise would not have been produced in this country at all and discouraged production in export industries.

The direct control of production by government authority is, however, vastly different from the intermittent and sometimes temporary government control in a price economy. When directed by prices, production is necessarily geared to consumers' wants in accordance with their incomes; when directed by government authority, as in a planned economy, production is geared to the purposes of the authorities. The purposes of the authorities may be the waging of war, preparation for war, intense industrialization, national aggrandizement, or consumer welfare as interpreted by the authorities, with or without the expression of the consumers themselves. Guiding of production by authority requires elaborate administrative apparatus, because all of the impersonal decisions made by the markets in a price economy must here be made by boards, commissions, committees, commissariats, bureaus, and similar bodies.

Still another alternative has been suggested: that production can be guided by authority through a process of "imitating" the system of free prices; prices would be set on a trial-and-error basis until the results of a "free price" system would be achieved. Prices would be so fixed by government that the amount of each product demanded would equal the amount

produced. While such a system has never been tried, wartime experience with price control raises grave doubts as to whether such price changes could be made fast enough and accurately enough to prevent serious distortions of the economic system.

Much can be said of the problems of incentives and management in an economy where government plans and carries out all production. But the great and all-important difference between production which follows prices and production which follows authority is that the former is a system under which individuals freely adapt their productive resources to meeting the freely made consumers' choices, while the latter is a system in which government directs production to its own ends.

Prices Determine Income

It has been stated that incomes depend upon prices. Wage and salary rates are prices paid for personal services. Rents and royalties are the prices paid for the use of natural resources and capital. In general, incomes are of two broad types: personal-service incomes, arising from the exertion of human effort, and property incomes, arising from the sale of rights to employ property (producer goods and rights to producer goods).

Incomes not only reward those who engage in production, they also serve to allocate productive services for different uses. Consider the allocation of labor. Why do young people choose particular occupations and professions? Working conditions, the kind of training required, the chances of advancement, social standing, individual likes and dislikes are all factors influencing the choice of a business, occupation, or profession. But a major and omnipresent factor is prospective earnings. Prospective earnings of the different occupations vary widely. In general people seek those with the highest incomes. Thus, income differences dominate in directing the flow of young people from the schools into the various trades and professions. It is recognized that the higher respect accorded skilled occupations makes them for some people more desirable than unskilled jobs, but the higher incomes of skilled workers constitute a more powerful attraction. Similarly, clerical jobs are often preferred over manual jobs paying somewhat more money because of the prestige accorded office work, yet when manual jobs pay markedly increased incomes—as during the war—many white-collar workers forsake their assumed dignity for fatter pay envelopes.

The price of labor, then, plays a major part in distributing labor among different employments, much as the price of a commodity distributes it among various uses. Changes in the earnings of labor are especially im-

portant. When earnings in one occupation or profession rise, without a corresponding rise in other fields, labor crowds into the higher income field. If carpenters' earnings in working on the construction of apartment houses become higher than carpenters' earnings in working on single houses, carpenters will seek employment on apartment projects. If freshmen engineering students believe that the incomes of chemical engineers will be higher than those of mechanical engineers, more of them will decide to become chemical engineers.

CRITICISMS OF THE PRICE SYSTEM

Prices, along with some government control, do the work of distributing goods, directing production, and allocating resources. What are some of the criticisms of the price system? One objection often voiced in the past is based on ethical grounds. In a price economy men are bound together by a cash nexus, it is declared. They work together not for reasons of brotherhood and humanity but because the acquisition of money is their common and overwhelming desire. The morals of a price economy are the morals of the market place: to buy cheap and sell dear, to be selfish and deceitful to advantage, to separate nobility and goodness from the work-a-day world. This type of objection is in part simply a criticism of all worldly institutions. Beyond this, however, it does bear specifically on money, prices, and profits as means of conducting economic activity. To be sure, money earnings are the reward and the economic measure of success for the farmer, the businessman, the salaried employee, and the professional man. But the money earnings of one person come from other individuals whom he has somehow served and who voluntarily decide to pay him for services rendered. The price system is a system of human co-operation, albeit an impersonal one which achieves the closest adaptation possible of resources to the ends of individuals whatever these ends may be.

Still another type of criticism of the price system is that it is not humanitarian or, at worst, degrading. Consider the plight of the hand-loom weavers during the early part of the nineteenth century. Textiles were formerly produced by hand, and the hand-loom weavers earned a subsistence income. The introduction of machinery during the Industrial Age radically changed methods of production in the industry, and the prices of textile products fell. The incomes of the hand-loom weavers were cut very sharply, so that in parts of northwestern Europe they suffered acute distress and many starved to death. The contemporary economy had no provision for the succor of its workers. This condition has characterized most societies

under a price economy at intervals during modern times. Individuals have been victims of the economy's changing methods and general fluctuations, as well as being benefactors of it. What is involved here are those instances where changes in prices drastically reduce the income of people who depend upon the sale of some one product or service for their livelihood, and who have no alternative source of income. When there are other income sources, the impact of the drop in one price is less severely felt. Yet any society, unless it is one in which no change of any kind takes place, is faced with the employment problem of those members whose services suddenly lose their value. The obvious thing to do about a situation of this kind is to direct such persons into branches of production where their services will have a value. But in a price economy, where the price and market mechanism is relied upon, individuals whose services decline in value are left to shift for themselves. They are given advice, unemployment compensation, and the benefits of some kind of social security relief but they are not actually transferred, by governmental authority, to other areas, industries, or occupations. This is the inevitable accompaniment of the freedom of choice afforded the individual.

A further criticism of the operation of the price system is technological, as it were, declaring that the vast and complicated production of modern industry is devoted, not to what is considered humanitarian and useful, but rather to what promoters and speculators guess will sell at a profit. Production serves pecuniary ends, not technological ones. This criticism has been of recent times particularly forceful in the United States where its slogan is often given as "production for use, not for profit." But from the price economy's point of view, production for profit is production for use. A producer makes profits only if people who want them will buy his wares; profits and prices are thus the link between the producer and the consumer, the mechanism which gears production to consumption. Critics of the price system often retort that producers drum up a demand for useless things and so make a profit for selling useless or even harmful goods, but this position inevitably is based on the assumption that consumers lack rationality and education to achieve adequate protection in the economy.

It is, furthermore, often stated that the price system is "unplanned," that it is anarchic, chaotic. It is admitted that the modern economy operates largely according to consumer preferences, but certain critics point out that the pattern of operation is unstable and is easily disrupted by depressions. Conceivably, stability could be attained in a rigidly planned economy. To plan an entire economy means, however, to substitute the decisions of government for those of consumers which would mean the sacrifice of

individual rights. A major problem is posed in modern times by those who advance the rival claims to stability and security by government control on the one hand, and to the retention of individual freedom through the price system on the other hand.

The price system adjusts production to the preferences of the consumers, but it is also argued that while wealthy consumers can satisfy their varied wants, poor consumers are often unable to satisfy very basic wants. While inequality of incomes is attributable to the operation of the price system itself, the extent of inequality is not inherent in the price system. Some institutions, such as inheritance, and the existence of unequal educational opportunities, greatly intensify the inequality arising under the price system.

Economists generally agree with another type of criticism—that based on market imperfections. By "market imperfections" are meant the impediments to the functioning of prices in allocating resources. Lack of knowledge is important; if the consumer does not know where to obtain the lowest prices, he fails to obtain maximum satisfaction from his income. Similarly, producers often lack knowledge of available market opportunities and thus obtain insufficient income from the use of their resources. Thus, where knowledge is poor, the adaptation of production to consumption is imperfect.

Another aspect of imperfection is the slowness with which adjustments to price changes are sometimes made. The price of a product may decline, thus indicating the need for a contraction in the output of the commodity and a consequent smaller number of producers, but since the contraction may occur over several years and meanwhile be interrupted by several periods of renewed high prices, producers are unable to adjust efficiently their inventories and schedules. Industries which are subject to a long-term decline in the demand for their products and therefore to long-term expectations of lower prices, but which decline slowly owing to periods of temporarily higher prices are known as "sick industries."

A further and related distortion of the price system is that exercised by monopolistic influences. For present purposes, monopolistic practices may be defined as control or manipulation of prices by persons or organizations. The resource-allocation function of prices is exercised most efficiently when prices are not influenced by the direct decision of any one person or group. This may be restated: the price system works best when prices reflect the sum total of the decisions of many individuals. Prices should determine, and not be determined; prices should direct individuals, not be directed by them. In wide areas of the present economy prices are in fact controlled by monopolistic organizations, by individual businessmen, by unions, and by the government. Often this control is minor; a manufacturer may be able to

decide to price his product at $9.90, rather than at $9.95, not being in a competitive position to deviate more than 0.2 per cent from a price of $10. In other cases the extent of the control is much greater. If an organization can control the price of an important raw material, a fairly substantial diversion of resources is effected.

CHAPTER 13

Demand

"Prices are determined by demand and supply." Editorials in the press and letters to the editors, like most public discussions of this day, have at one time or another declared that some measure "violates the law of supply and demand." The "law" of supply and demand is invested thus with the dignity and the inviolability of the law of gravity. Unwary indeed must be the legislator who would dare go against such a law. There is, of course, no unique law of demand and supply in the sense of some rigid relationship which always manifests itself in the same way. Writers on economics in the nineteenth century did lay down "laws of political economy," and did make some pretensions that these were similar to natural laws, but a preoccupation with economic laws is not a fruitful pursuit in economics today. There is quite obviously a need to state conclusions about, for example, the relations between demand and price under certain conditions; but nothing is gained by calling such conclusions laws. The careful student will arrive at a series of generalizations, bearing one on the other, about demand and supply, not one.

The prices of commodities and services, it has been said, are the rates in money at which the commodities and services are exchanged for one another. In each of the millions of exchanges taking place every day there is a buyer who pays a price for the commodity, and a seller who sells the commodity at a price. The following discussion is centered on the action of the buyer, the reasons for it, its characteristic behavior, and its consequences. This is the problem of analyzing demand, which is one phase of the inquiry into the forces determining prices.

AGGREGATE AND PARTICULAR DEMAND AND SUPPLY

At this point an important contrast should be noted between the description of demand and supply in Chapters 4-6 and the definitions that are employed in this and following chapters. In Chapters 4-6 the term "aggregate demand" represented the total expenditure of consumers, investors, and government for commodities and services. In the present discussion, however, the analysis of demand focuses on the *individual* demand for a specific product and, likewise, on the demands of all persons in a given market for a specific product. Aggregate demand, on the other hand, refers to the whole expenditure for the national output. This contrast is evident also in the definition of supply. Unlike *aggregate* supply, which is the total value of the commodities and services making up national output, the analysis of supply in the following chapter focuses on the supply of *particular* commodities and services and the conditions governing this supply. The two approaches—the aggregate and the particular—are often distinguished as the *macrocosmic* and *microcosmic*. The macrocosmic is concerned with the whole, the microcosmic with the parts comprising the whole. For purposes of analyzing the levels of national output, income, and employment the aggregative concepts are used; for analyzing the composition of the national output the particular demand-and-supply approach is used.

PATTERNS OF CONSUMER EXPENDITURE

Granting that the spending decisions of individuals and families direct the employment of resources and influence the use of these resources, it is important then to examine the effects of these decisions on the prices of particular commodities and services. How American consumers divided their expenditures in 1935-36 is shown in Table 23.

Of the total expenditures, approximately 75 per cent was spent on food, clothing, and shelter. It is particularly important to compare the expenditures of persons in relation to their individual incomes. Consumers with annual incomes between $750 and $1,000 devoted 40 per cent of their expenditures to food, those with incomes between $1,750 and $2,000 spent 33 per cent on food, and those with incomes between $10,000 and $15,000 spent 20 per cent. Although the high-income consumers spent only half as large a proportion of their incomes on food as the low-income consumers, the high-income consumers spent on the average over six times as much on food as the low-income consumers. For the other items of expenditure,

Table 23 **DISTRIBUTION OF CONSUMER EXPENDITURE, 1935–36**

Expenditure	*Average percentage of individual income*
Food	33.6
Housing and household operation	29.5
Clothing	10.5
Automobile	7.6
Medical care	4.4
Recreation	3.3
Furnishings	2.8
Personal care	2.0
Tobacco	1.9
Transportation other than auto	1.7
Reading, recreation and other expenditure	2.7

Source: Adapted from National Resources Committee, *Consumer Expenditures in the United States* (United States Government Printing Office, Washington, 1939), p. 46. The term *consumers* included all family members (115,966,000 persons) and all single individuals (10,058,000 persons), but excluded residents in institutional groups (2,000,000 persons).

other differences, though less striking, will be found between low-, middle-, and high-income groups. In general, as the individual level of expenditure increases, the proportion of expenditures on food decreases; on housing it remains almost constant; on clothing, automobiles, medical care, recreation, and education it increases. For transportation by other means than by automobile the proportion declines at first and then rises with total expenditure. This variation is probably explained by the fact that the low-income groups are dependent on streetcar or bus transportation, the cost of which takes a larger share of their expenses than of those of consumers with somewhat higher incomes, while at the other end of the income scale, the well-to-do and wealthy generally travel extensively enough by rail, ship, and air to cause a larger share of their expenses to go for such purposes as compared with the middle-income groups.

The influence of income differences on the expenditures of whole groups of the population has been suggested only briefly. Income variations are of course not the only cause of differences in spending patterns; individual habits also have a bearing on spending patterns.

Problems of the Consumer

The consumer "problem" has been given increasing attention in recent years. The present-day consumer is regarded as a person who needs pro-

tection and assistance in his buying; laws are directed toward his safety and organizations and publications seek to advise him. One aspect of the consumer problem is the wide opportunity for fraud and misrepresentation by sellers which modern industrial society offers. There is nothing new about buyers being cheated by sellers, for fraud is as old as trade itself. The legal maxim *caveat emptor*—"let the buyer beware"—has long stood among people who prided themselves on being sharp traders. But in modern urban society the buyer is the object of much solicitude. Furthermore specialization of labor and the vast variety of products have led to a condition in which most people are not able to judge competently the qualities of much they buy.

First of all, there is the problem of the consumer's lack of knowledge. The present-day consumer must perhaps choose two hundred items from among many thousand, but his decisions are complicated by the fact that the two hundred commodities may appear under the guise of two thousand brands. How is the consumer to know which brand is best for his purpose? Advertising, which is in a sense also an adviser to the consumer, is not usually very helpful in the comparison of commodities. In fact, it is difficult to imagine anyone weighing as carefully as knowledge permits the merits of each of many brands of a commodity. Most consumers buy from habit, or on the advice of friends, or haphazardly, or at the behest of the loudest advertiser.

To be a competent producer each person is in this day a specialist, even if "only" in running a household, but to be an equally competent consumer each person would almost have to be a dietitian, a chemist, a metallurgist, a mechanical engineer, a physician, and a host of other specialists combined. Most of the buyers of automobiles or house heating systems or radios have no more than a superficial knowledge of these products.

Two aspects of the lack of consumer knowledge must be distinguished. One aspect is the consumer's carelessness in making expenditures. Most people are careless about some of their expenditures; one pays 10 cents for an article when it could be bought a block away for 8 cents; still another overlooks an auto dealer who would have offered him 50 dollars more on the old car. But carelessness, which is defined here as inattention to comparison of prices, is an individual matter. So long as individuals differ from one another, some persons will get more for their money than others do. The second aspect of the lack of consumer knowledge is one person's judgment on another's action as a consumer. One person censures another because he tries to cure a physical ailment with a patent medicine which is known to the first as worthless. One is annoyed by the other's ignorance. Ignorance in this sense is universal. Since human beings differ in habits,

tastes, education, training, and intelligence, there is a great deal of "ignorance" or behavior which is disapproved. Thus, the various drives for "consumer education" may represent in good measure the imposition of a set of values or judgments by one group on another.

The second major consumer problem arises from the loose relation between price and quality of commodities. The consumer must often make a choice between the better quality at a higher price and the poorer quality at a lower price. However, price and quality (in an objective sense) are not always equated in some products (such as drugs); better quality may sell at a lower price. Furthermore, the relation between prices and quality is also complicated by the problem of what constitutes quality; the consumer may be puzzled as to whether the difference in quality is actually measured by the higher price. The de luxe model of an automobile may or may not be "worth" its additional cost.

A third major problem is the consumer's reaction to advertising. To what extent are individuals' wants autonomous and to what extent are they derived from the influences of advertising? Advertising is used here to include not only printed matter, billboards, and radio, but also window and counter displays, the design and layout of stores, and the technique of selling in all forms. Advertising actually does create many wants, although perhaps not so greatly as advertisers claim.

There are considerable areas where the advertisers encounter much resistance. It is probable that no amount of advertising could induce men to change the styles of their clothes as often and as drastically as women do. No manufacturer succeeded during the 1920's and 1930's in widely marketing a very small passenger car despite big advertising expenditures. The effect of advertising is difficult to measure, especially when one considers that many commodities are bought in large amounts without advertising, and that generations ago consumers decided what to buy with few or no advertisements to guide them.

Fourth, there is the problem of the consumer's cultural heritage and background. Each human being has his own quirks and foibles, but he is largely influenced by the forces of religion, schooling, geographic region, social class, occupation, clubs, associations, and societies. Religion may dictate to the individual the consumption or the avoidance of certain foodstuffs. (Orthodox Judaism prohibits pork, Roman Catholicism prohibits meat on Friday, many Protestant denominations are opposed to the use of spirits and tobacco, etc.) Schooling may form tastes in clothing, recreation, and in the choice of possessions, such as books. Geographic factors obviously influence clothing preferences. And social class or occupation may determine

the whole pattern of consumption, since most persons are impelled somewhat by the desire to live in accord with a group of like means, or social standing.

These problems of the consumer form a background for the analysis of demand. It has often been said that a logical analysis, or theory, of demand is not possible because it must be predicated on the behavior of the individual human being who, it is argued, does not act consistently according to reason. But the concern here must be not with what the consumer should want, but rather with what he does want. An analysis of demand cannot assume that consumers behave rationally in all instances; rather, it accepts the conclusion that consumers know what they want, for whatever reasons, and that they usually try to better their material well-being.

THE THEORY OF INDIVIDUAL DEMAND

The theory of individual demand, or, as it is sometimes known, the "theory of consumption," is a formal explanation in the most general terms of the forces which determine the demand of an individual. This theory is important because the demands of individuals constitute market demand, and market demand influences prices, which in turn direct economic activity. To understand both how and why prices direct production and consumption, it is necessary then to explore the demand behavior of individuals. Demand analysis is vital, too, in explaining consumer expenditures which constitute the bulk of aggregate demand and therefore affect the level of national output. The behavior of the individual in demanding goods and services, then, influences both the allocation of resources as well as the volume of national output.

The demand of an individual for a commodity is based on three factors: (1) *wants,* or *preferences;* (2) *prices;* (3) *income.* The individual's demand for goods and services will be shown to depend on his wants or preferences (likes and dislikes), on the prices which face him, and on the size of his income.

Wants and Preferences

Everyone is conscious of his wants and desires, of his likes and dislikes, and it is obvious that these in some way direct his purchases of goods. The most workable explanation of individual wants in this sense is that a good is wanted because it possesses *utility. Utility is want-satisfying power.* In this definition, utility is not identical with usefulness, because the latter

implies the judgment, whether based on moral or scientific or traditional causes, of another person or a group on the individual. A man may be scorned by his friends because he carries about with him a gadget which never "works." The gadget is said to have no usefulness. But the owner treasures it, to him it has want-satisfying power; it therefore possesses utility. The vegetarian denies the usefulness of meat, but if an economist were himself a vegetarian, he would admit that meat has utility because meat satisfies the wants of some people. Thus, any article wanted by anyone for any reason possesses utility.

Next consider the relation between the quantity of a commodity wanted by an individual and its utility to him. As an illustration, assume that a pair of dress shoes has a certain utility to him. If Mr. A had two pairs of shoes—identical ones—the utility of both pairs together would be higher than that of one pair alone, i.e., he is better off with two pairs because he can, for instance, wear one pair while the other is drying after a rain. Further, if Mr. A had three identical pairs of dress shoes, the total utility of the three is greater than the total utility of two pairs. But if he had four pairs, he might not consider himself better off than if he had three. A fourth pair might indeed crowd his closet and become a nuisance rather than a benefit. The total utility of four pairs would be the same as, possibly even less than, the total utility of three pairs. As a person considers increasing the stock of a commodity, *the total utility of the stock to him will increase with the quantity up to a point.* But the increase in total utility to Mr. A from two to three pairs is not so great as the increase from one to two pairs. *The addition to total utility resulting from the addition of another unit of the commodity to the stock is its marginal utility.* In the foregoing illustration, the marginal utility of a pair of dress shoes to Mr. A diminishes as the number of pairs increases to the point that the marginal utility of the fourth pair is zero or negative. He does not want a fourth pair, because with three pairs he has all the dress shoes he wants. The foregoing illustrates the *principle of diminishing marginal utility: the utility of a unit of a commodity to an individual diminishes with increases in the number of units he considers.*

The relation between utility and quantity is given graphic illustration in Chart 6.[1] In part *A* of this chart the bars represent the *total* utilities of each quantity. As quantity increases, the heights of the bars also increase. The heavy line drawn through the mid-points of the bars can be called the

[1] It must be understood that utility per se is not measurable. The vertical scales in Chart 6 represent only *more or less*, not specific quantities. See also pp. 296-98 for the scale of preference concept which avoids use of utility concepts.

curve of increasing total utility. It is to be noticed that the curve rises at a diminishing rate: i.e., it becomes flatter. Part *B* of Chart 6 shows the curve of diminishing marginal utility. This curve is drawn through the mid-points of bars which show the marginal, i.e., additional utility. It can be seen that the bars in part *B* are taken from those in part *A*: the shaded areas in part *A* are the increases in total utility as quantity increases, and the same increases are shown separately in part *B*.

The principle of diminishing marginal utility is general in definition and of the widest application. When correctly applied it is valid for all individuals and all commodities; it is, however, frequently misunderstood and misapplied. Correct application of the principle depends on these elements:

1. The principle must be applied to a particular commodity, not to a general class of commodity. In the illustration above, *dress* shoes were used. The classification shoes as such would not serve the purpose of this illustration because there are many types of shoes: street, sport, golf, dancing, etc. A golf shoe is a different commodity from a dress shoe and in careful analysis must be treated separately.
2. The principle is valid only for an interval of time during which an individual's want for a commodity remains unchanged.
3. The principle applies not to the "last" in a series of identical commodities, but to *any one* in the series. The fourth pair of shoes to Mr. A had a marginal utility of zero or perhaps a negative utility. But the fourth pair is itself blameless. Since the four are all alike, any one of the four has the same marginal utility.

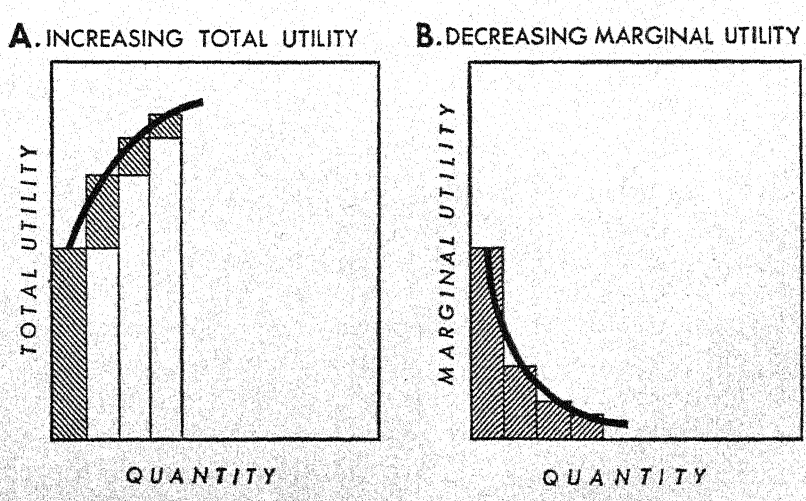
Why does marginal utility diminish as quantity increases? Satiety is sometimes given as a reason, but this is not helpful. Still another reason states that the responses to a repeated stimulus tend to weaken successively. A simpler and sounder explanation of diminishing marginal utility is this: successive units of an identical commodity are applied to less and less important uses to the individual. A second pair of shoes serves as a reserve pair and is thus less important than the first pair. Eggs, for example, can serve a variety of culinary ends; the utility of one dozen eggs thus varies according to the weekly consumption, because when there are more dozens, the additional eggs are put to less important uses.

The diminution of marginal utility varies from commodity to commodity for any one person. The curve of diminishing marginal utility (see part *B* of Chart 6) can thus have different rates of decline for different commodities. During a year, an individual may buy many tons of coal, hundreds of quarts of milk, several shirts, but only one hat (of a given kind); in some

of these commodities the utility of additional units diminishes rapidly, in other cases less rapidly. Of some articles, like automobiles, one unit is bought only once every few years; most people who buy houses buy only one in a lifetime. Watches are bought one at a time and usually at very long intervals. For commodities such as these the principle still applies: the second unit has a very low utility or perhaps none at all.

So far this analysis has dealt with the desire of an individual for *one* commodity at a time. The principle of diminishing marginal utility can also be used to describe the relations between wants for several commodities. For indeed each individual has a structure of wants. Assume a case in which price and income have no influence: e.g., a man embarking upon a long hunting trip who is permitted to take with him whatever he selects from a store of hunters' and trappers' provisions. The hunter will be guided in his selection by (1) the weight of the things he can carry, and (2) the utility of the various provisions. He takes one frying pan, although he could easily use two when cooking. He takes 20 pounds of flour. The utility of the nineteenth and the twentieth pounds of flour exceeds the utility of a second frying pan (assuming that it weighs 2 pounds). He takes 3 pounds of coffee, the marginal utility of which is greater than the utility of a second frying pan. If he hesitates as to whether to take 21 pounds of flour and 2 pounds of coffee, or 20 pounds of flour and 3 pounds of coffee, then with these quantities the marginal utility of flour to him is about equal to the marginal utility of coffee. If the total weight of the hunter's pack is 60

Chart 6 **THE RELATION BETWEEN UTILITY AND QUANTITY**



pounds, and if he makes his plans very carefully, the marginal utility of a pound of one commodity will be about equal to the utility of a pound of any other commodity. Where the units of commodities are the same—in this case pounds to be carried—the individual chooses different quantities of the commodities he wants and which are available to him in such a manner that the marginal utilities of all the commodities are equal.

The principle that the individual chooses different quantities of equally measured commodities in such a way that their marginal utilities are equal, has general applicability. The customary method of measuring commodities in a market is price. When commodities of equal price are available the individual chooses among them in such a way as to equalize the marginal utility of each. However, because prices of commodities are not equal it will be necessary to explain later how the individual arrives at a composition of purchases which maximizes the utility derived from these purchases. Before this is explained another approach to the demand of an individual for commodities will be examined.

The Scale of Consumer Preferences

Up to this point the analysis of individual demand has been built upon the idea, or the concept of, utility. An objection often stated against this concept is that utility in itself is not directly measurable. If a smoker gets more utility from a pack of cigarettes than from a cigar, the question arises: how much more?

One possible answer is to say that a commodity has a utility of a certain number of dollars and cents. But there is a fallacy in money measurement since money means different things to different people and to the same person at different times. The marginal utility of money is itself variable, for the utility of one dollar depends on how many dollars a person has. Thus money is not an invariable standard.

Another objection often made to the utility concept is its supposedly invalid psychology. It has been argued that the utility concept is tainted with association with the outmoded pain-and-pleasure psychology of the nineteenth century.

These two objections have been partly responsible for establishing an alternate approach based on the *scale of preferences* concept. Each individual has a scale of preferences for commodities and services. The individual says, "I know what I like." His likings can be extended to a list, or scale, of commodities and services—different amounts of each—in the order of

their importance to the individual. If one commodity has a higher place on the scale than another commodity, it need not be said that it has greater "utility" but simply that it is *preferred.* By placing the emphasis in this way, the whole problem of the measurability of utility is circumvented.

The preference-scale concept also employs the assumption that all commodities and services are *substitutes* for all others. The word "substitute" in economics does not connote inferiority. Beef is a substitute for veal which is a substitute for lamb which is a substitute for pork which is a substitute for beef. Thus, in this context substitution is the interchangeability of commodities in the budgets of consumers. Even physically different commodities may be considered substitutes for each other. Young married couples may hesitate between a new automobile and better housing; if this is so, then houses and cars are substitutes. A magazine is a substitute for a movie, a new dress for three pairs of new shoes, and so on through the whole gamut of commodities and of wants and preferences.

The concept of a scale of preferences permits the conclusion that in an objective economic view there can be no classification of goods into luxuries and necessities. Cigarettes may be high on the scales of preferences of some persons, yet have no place at all on the scales of preferences of others. Milk is both a necessity, i.e., high on the scale to some, and a luxury, i.e., low on the scale to others. A's necessity is B's luxury and vice versa. By common definition luxuries and necessities are usually always invested with the meanings from moralistic or political judgments, but in the preference-scale concept these judgments have no part, for this concept simply states that goods vary in importance to different individuals.

An individual's scale of preferences is of course not a static thing. It changes as the individual learns of new commodities and of new ways of consuming old ones. It changes as frequently as does the totality of the forces which determine an individual's wants. A common experience is for the preference scale to change as income rises. Of course, a higher income may result in consumption according to an unchanged preference scale. Just because someone consumes more at a higher level of income does not necessarily imply a new scale of preferences; it may signify only that more wants can be satisfied.

These two approaches to the theory of individual demand—through marginal utility and through the scale of preference—are not antithetical to each other. Their important differences are chiefly matters of emphasis. To portray the way in which wants are transformed into demand, it is simpler

to use the marginal-utility concept. But after all, the utilities of different goods to an individual can be interpreted as a matter of a scale of preference; and the hierarchy or scale of preference of an individual rests upon his estimate of the varying utilities to him of different goods and services.

The Role of Prices in Determining Consumption

Only when expressed in money expenditure do wants or preferences influence prices. The expression "effective demand" is sometimes used to give emphasis to this notion. Wants exert their force on the stock of *available* goods—those already produced and those planned to be produced. The mere physical existence of commodities is not the most important aspect of availability; far more important are the prices at which commodities are available.

The different prices of commodities indicate the rates at which, taken one with another, they are available to the individual. When faced with the choice of two goods of equal marginal utility, or concerning which he has no special preference, the consumer tends to buy the cheaper. The general method in economics is to distinguish clearly and sharply between the effects of price-determined and preference-determined influences on consumption; purchases during a period show a person's tastes or preferences, but his purchases also show the effects of prices.

The demand of an individual for one commodity is expressed as a *demand schedule: the amounts of a commodity that would be purchased at a given time at different possible prices.* An individual demand schedule is drawn upon these assumptions: (1) an individual's wants are given, (2) his income is given, and (3) the prices of all other commodities are also given. A schedule representing the hypothetical demand of an individual for commodity X is shown in Table 24.

Table 24 **DEMAND SCHEDULE OF AN INDIVIDUAL**

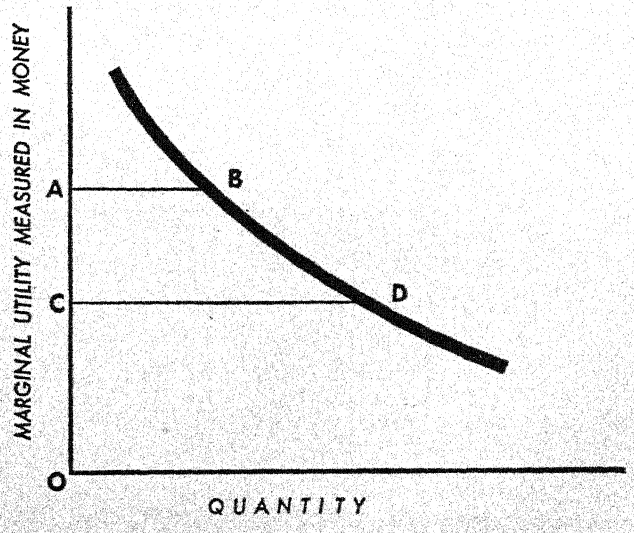
Possible prices of commodity X	*Number of units which would be purchased*
$2.00	1
1.00	2
.75	3
.60	4
.50	5
.40	6
.20	7

The schedule shows that if the individual must pay $2 for a unit of X he would buy only one unit; if the price were 75 cents, he would buy three units, and so on. If the price actually is 50 cents, he will buy five units; he would not buy six units of X, because a sixth unit is worth only 40 cents to him. He would not buy only four, because he would then forego the opportunity of getting a unit of X for 50 cents when it is worth 50 cents to him.

Chart 7 shows the *demand curve* of an individual. This is the graphic representation of an individual's demand schedule for a commodity or service. In the chart the horizontal scale shows quantity, while the vertical scale measures, in money, the marginal utility of the commodity to the individual. Thus, if the price of the commodity is *OA* the individual buys an amount *AB*, while if the price is *OC* he would instead buy the amount *CD*. The fact that the demand curve slopes downward from left to right is determined by the principle of diminishing marginal utility. The position of the curve indicates (1) the intensity of the individual's want for the commodity at any possible prices, and (2) the amount of his income.

It is possible to construct a *demand schedule* for any individual, based on every commodity that has utility for him. In some instances if the price of the good is higher than the marginal utility expressed in money, the individual will buy none; in other instances, one or more units of a commodity will be bought at the price the consumer must pay. *The consumer*

Chart 7 **THE DEMAND CURVE OF AN INDIVIDUAL**



seeking to get the most from his money expenditures[2] *will so direct them that the marginal utility of each commodity bought bears the same proportion to its price as the marginal utility of each other commodity bears to its price.* Thus the ratios of marginal utilities are equal to the ratios of prices. This principle is significant because it sums up the theory of individual demand and because it explains the apportionment of expenditures by an individual among different commodities.

As an illustration of the principle, consider the purchases of a veteran returning to civilian life with back pay and allowances amounting to $1,000.00. For five of the commodities (of an assumed quality) desired by the veteran the demand schedules are shown in Table 25.

Table 25 DEMAND SCHEDULE OF A VETERAN

(Possible prices and number of units he would buy at each price)

Automobiles		Suits		Shoes		Hats		Shirts	
$800	1	$80	1	$20	1	$10	1	$10	1
200	2	50	2	10	2	4	2	6	2
		40	3	5	3			5	4
		30	4					4	6
								3	10

If the prices prevailing are as follows: the automobile $800, a suit $50, a pair of shoes $10, a hat $10, and a shirt $3, the veteran will buy one automobile, two suits, two pairs of shoes, one hat, and ten shirts. This distribution of his expenditures will maximize the utility from the money he spends on these five commodities. For example, if he bought only one suit instead of two, the suit would be worth $\frac{1}{10}$ of an automobile to him, but would cost him only $\frac{1}{16}$ as much. Therefore, he gains in satisfaction by buying two suits. When he does so a suit will be worth $\frac{1}{16}$ of an automobile to him, and will cost him $\frac{1}{16}$ as much. If he bought three suits, each would be worth to him $\frac{1}{20}$ as much as an automobile but would cost him $\frac{1}{16}$ as much. Obviously, he would not buy a third suit. The same relationships hold true for any other pair of commodities. Thus satisfaction of an individual is maximized when the ratios of the prices of the commodities bought are equal to the ratios of their marginal utilities.

[2] This is not the same as getting the most from his *income,* for some part of income may be *saved.* He will distribute his income between consumption and saving so as to derive an equal marginal utility from each dollar however disposed.

The Role of Income in Determining Individual Demand

The third major determinant of the demand of an individual is income. Individual purchasing depends on income as well as upon wants and prices. The larger the income of an individual the more of his wants he can satisfy. As his money income rises, the marginal utility of money—the utility of one dollar—diminishes; thus as his income rises the consumer's demand schedules for separate commodities alter. If the preceding schedules (Tables 24, 25) applied to a larger individual income, the figures under the heading "Number of units which would be purchased" would all increase, signifying greater purchases by the individual. As income rises, purchases of different commodities do not necessarily increase correspondingly. In fact, it is possible that the purchases of some commodities fall off as income goes up. The reason is that as an individual expands his purchases, the importance of the various goods in his preference scale shifts. A poor family consumes margarine; but if its income increases somewhat it will consume more margarine and finally, perhaps, it will change to butter.

MARKET DEMAND

The individual demand schedules for one commodity, when added together, form a *market demand schedule* for that commodity. A *market demand schedule shows the amounts of a given commodity that would be bought at different possible prices in a given market area at a particular time.*

The market demand schedule is presented in the same way as the individual demand schedule. Amounts of commodities which buyers collectively are willing to purchase are compared at various prices in a given time period. A hypothetical market demand (in a small town on some one day) for sirloin steak is given in Table 26.

Table 26 **MARKET DEMAND SCHEDULE**

Possible prices of commodity	*Number of units which would be purchased*
$0.80	500 lbs.
.75	600
.70	800
.65	1,000
.60	1,200

The prices are *possible* prices, and they are *simultaneously considered.* The price of 80 cents is not established before the price of 75 cents, and so on; on the contrary, each price is considered as being a possible charge at the same time. Thus, *if* the price were 80 cents, *then* 500 pounds would be bought; *if*, on the other hand, the price were 75 cents, then 600 pounds would be bought. Demand is not an amount standing by itself but always a series of amounts associated with a series of prices—or, demand is a schedule of amounts associated with a schedule of prices. It is almost meaningless to say, for example, "There is a demand for 2 million houses," because the prices of houses are not considered therein.

The relationship between amount bought and price in a market demand schedule is an inverse one, that is, at a lower price a large amount would be bought, and at a higher price a small amount would be bought. This relationship, which is sometimes referred to as "the law of demand," is explained by (1) diminishing marginal utility, (2) differences in scales of preference, and (3) differences in individual incomes. In the first place, the factor of the diminishing marginal utility of the commodity to the individual plays a role because each buyer in the market will buy more at lower prices, or, alternatively, the price must be lower to induce him to buy more. This amounts only to saying that a market demand schedule is a composite of individual demand schedules, and reflects the principal characteristic of the individual demand schedules. A second reason for the inverse price-demand relation follows from the differences among individual demand schedules, which in turn arise from differences in scales of preference. Some individual demand schedules begin with one unit bought at a high price, while in others the first unit is bought only at a lower price. Hence, at successively lower prices new demand schedules of individuals come in for the first time. Third, income differences are very important. Generally speaking, at high prices the buyers are mainly the well-to-do, and since there are comparatively few with high incomes, the amount bought at high prices is small. At successively lower prices more is bought, because the middle and lower income consumers enter the picture as buyers.

Changes in Market Demand

Next, consider the meaning of increases and decreases in demand. Changes in demand are changes in the entire schedule. Consider Table 27. Here are three hypothetical market demand schedules, A, B, and C. Schedule A will be recognized as the same as that appearing above in Table 26.

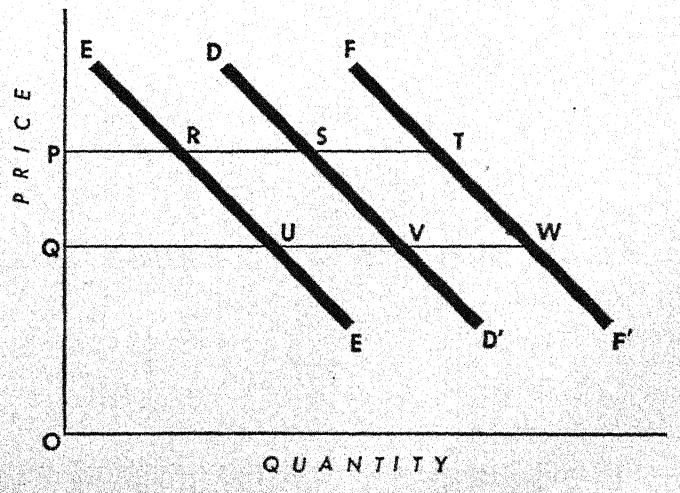
Table 27 CHANGES IN MARKET DEMAND

Schedule A		Schedule B		Schedule C	
Possible prices of commodity	*Number of units bought*	*Possible prices of commodity*	*Number of units bought*	*Possible prices of commodity*	*Number of units bought*
$0.80	500 lbs.	$0.80	600 lbs.	$0.80	400 lbs.
.75	600	.75	800	.75	550
.70	800	.70	1,100	.75	700
.65	1,000	.65	1,400	.65	900
.60	1,200	.60	1,700	.60	1,050

Demand schedules B and C are different demand schedules for sirloin steak in the same small town, but at different times from that in schedule A. In schedule B, demand is greater than in A, greater at each of the prices listed; it represents an increase in demand. *An increase in demand means that buyers take more at all prices, and that at any one price the amount which would be bought is larger.* Schedule C shows a smaller demand than A; here there has been a decrease in demand from A to C. *A decrease in demand means that buyers take less at all prices, and that at any one price the amount which would be bought is smaller.*

Chart 8 shows changes in demand. The line DD' represents the demand for some commodity at one point of time; at the price OP the amount PS is bought, and at the price OQ, the amount QV. The line FF' represents an increase in demand over that shown by DD'; at prices OP and

Chart 8 CHANGES IN DEMAND



OQ, larger amounts, namely *PT* and *QW*, are bought. Similarly, the line *EE'* represents a decrease in demand from *DD'*. It is to be noted that while the demand "curves" in Chart 8 are drawn as straight lines, a demand curve may be of any shape, depending upon the particular commodity and the particular circumstances involved. A demand curve, however, slopes downward from left to right, whatever its shape may be.[3]

The causes of changes in demand are: (1) an increase or decrease in the incomes of the buyers; (2) a shift in the preferences of the buyers; and (3) an alteration in the price of a related commodity. Thus, a demand schedule will increase, as from schedule A to schedule B above, or a demand curve from *DD'* to *FF'* as in Chart 8, if any one or a combination of the following happens: the buyers' incomes rise, their wants for the commodity become more intense, and the price of a substitute commodity rises. Demand decreases when the three factors change in the opposite direction.

In the study of demand schedules it is necessary not to confuse the change in the relation between amounts bought at prices *within a single demand schedule*, on the one hand, with the change *between* schedules on the other. In any demand schedule, the amount bought is less at higher prices than at lower prices, but this has nothing to do with a decrease in demand. If a certain demand decreases, this means that a new schedule replaces an old one and, further, that in the new schedule all of the amounts for each of the prices are smaller. The demand-schedule concept is highly effective in dispelling inaccurate reasoning about demand and price relations.

Elasticity of Demand

Since commodities and services differ so widely from one another, it is to be expected that the relation between price and demand will vary from one demand schedule to another.

Table 28 DEMAND SCHEDULES OF DIFFERENT ELASTICITIES

Schedule A		*Schedule B*		*Schedule C*	
Possible prices of commodity	*Number of units bought*	*Possible prices of commodity*	*Number of units bought*	*Possible prices of commodity*	*Number of units bought*
$1.00	1,000	$1.00	1,000	$1.00	1,000
.80	1,100	.80	1,250	.80	2,000

[3] Except in rare cases which may be ignored here.

Table 28 gives the market demand schedules for three different commodities, A, B, and C. In all three cases the amount bought at a lower price is larger than at a higher price, but in A the proportionate change in the number of units bought due to the change in price is slighter than in B, and it is greater in C than in B.

The relation between proportionate change in price and proportionate change in the amount bought in a demand schedule is *elasticity of demand.* Elasticity means responsiveness of the quantity bought to a change in price. When a change in price of, for example, 20 per cent is accompanied by a change in quantity purchased of 10 per cent, the demand is said to be *inelastic.* This is illustrated in schedule A in Table 28: the change in price from $1 to 80 cents is a 20 per cent decrease, but the amount bought increases only from 1,000 units to 1,100 units, i.e., a rise of 10 per cent. In schedule C the opposite condition prevails: a 20 per cent drop in price is matched by a 100 per cent gain in the amount bought. When the increase in amount bought is larger proportionately than the diminution in price, demand is said to be *elastic.* Schedule B lies in between A and C. When price declines it falls to $\frac{4}{5}$ of its former level, while amount bought increases to $\frac{5}{4}$ of its former size. When the change in amount bought is proportional to change in price, the demand schedule is said to have an *elasticity of unity;* there is, in other words, a one-to-one relation between a fall in price and a rise in amount bought.[4] Any one demand schedule may exhibit varying degrees of elasticity at different prices.[5]

Chart 9 displays three demand curves. In each case, two prices are denoted on the graphs, namely *OP* and *OQ*. When the price changes from *OP* to *OQ*, or vice versa, the quantity bought also changes. In case A, that of a commodity with an inelastic demand, the graph shows a slight change in quantity, from *OA* to *OA'*. In case C, however, the change in quantity is relatively a large one: from *OC* to *OC'*; this is a commodity with an elastic demand. Case B, of the commodity with unit elasticity of demand, shows a change in quantity—from *OB* to *OB'* which is proportionate to the change in price from *OP* to *OQ*.

It is customary to illustrate elasticity by referring to observed price-quantity relationships; these illustrations must be understood to refer only to the responsiveness of quantities bought to changes in prices within a certain observed range. A favorite example of an inelastic demand is salt;

[4] The mathematical expression of elasticity is $-\frac{p}{q}\frac{dq}{dp}$ where p is price and q is quantity, and where the demand function is $q = f(p)$.

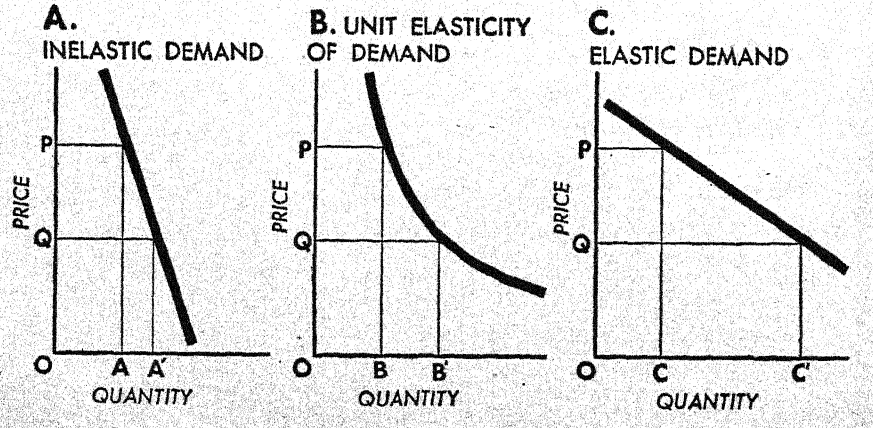
[5] See p. 308.

the demands for cigarettes and for gasoline are likewise inelastic. On the other hand, the demand for sirloin steak is probably elastic, as is the demand for automobiles and private airplanes. The demand for pork is thought to be of unit elasticity.

The elasticity of demand for a commodity depends on (1) the kind and number of substitutes for it, (2) the number of uses to which it can be put, and (3) the relation between its price and the incomes of its buyers.

Commodities for which there is a large number of close substitutes have a very elastic demand. Commodities which are produced in many brands ordinarily illustrate elastic demands. Although manufacturers may seek to convince consumers of the special and unique qualities of their own brands, each brand of a commodity is a close substitute for the other brands. If one producer of a leading brand of cigarettes lowered his price below his competitors, his sales would probably increase in larger proportion than his cut in price: the demand for the *one* brand of cigarettes is very elastic. In housebuilding materials there is a wide range of substitutes. Different kinds of lumber can be substituted for one another for many purposes; the same is true of many hardware items. The demand for any one of these materials is very elastic, for a slight rise in price for one material would divert a large part of the demand to other materials, and vice versa. Commodities for which there are no close substitutes are likely to have an inelastic demand. Despite a rise in price there is no close substitute for the consumer to change to and hence he will buy nearly as much as he did at a lower price. A fall in price does not increase the quantity purchased greatly because buyers are not attracted away from close

Chart 9 **ELASTICITY OF DEMAND**



substitutes. Tobacco has no close substitute and the demand for tobacco is therefore inelastic. There seems to be no close substitute for alcoholic liquor in its various forms, although reformers have spent much effort to increase the liking for substitute products or "interests."

The greater the number of uses to which a commodity can be put the more elastic the demand; the fewer the uses, the more inelastic the demand. Wheat is used as animal feed as well as human food; thus, the demand for wheat at lower prices is extremely large, because at this price farmers will use more of it to feed to their poultry and their animals. If automobiles were cheap enough, many families would buy a second car. Eggs can be put to a wide variety of uses in the kitchen; the demand for them is elastic, because the housewife will find various additional uses for eggs as their price becomes lower. The "lowest" use of lumber is as fuel, and if the prices for lumber fell far enough the amounts bought would increase greatly. On the other hand, there is not much one can do with cigarettes except to smoke them; they have no subsidiary uses such as feed, fodder, and ornamentation; and this is still another reason for their inelasticity of demand.

When the price of a commodity is low and when the monthly or annual expenditure on the commodity is a small proportion of income, demand tends to be inelastic; when there is a high price-income ratio, demand tends to be elastic. An American family spends only a few cents a month for salt, and the monthly expenditure on salt is a usually negligible fraction of the income of even a low-income family. If the price of salt rises, the effect on the family budget is scarcely appreciated and hence there is little reason to use salt sparingly. Similarly, if the price of salt falls, the purchase of it will hardly increase, since almost all consumers will already be buying as much as they want of it (owing to its low price in relation to their incomes). Only those who buy large amounts of salt, such as bakeries and restaurants, have any reason to watch their salt consumption and expenditures closely. On the other hand, the price of an automobile is a major item in the budgets of nearly all buyers so that a change in the average price of new automobiles has an important effect on the number bought. It is often said that the demand for luxuries is elastic, yet it has been indicated that the concept of luxury has no meaning except for a particular individual. Actually, "luxuries" are commonly the articles with the high price-income ratio, such as radios, foreign travel, summer houses, and automobiles. Expenditure for a good with a high price-income ratio normally entails the sacrifice of many other commodities of lower ratio. A rise in price sharply restricts amount bought, because it puts the good beyond the

reach of many buyers, since the added sacrifice of other commodities is more than they feel they can bear. A fall in price, on the other hand, stimulates amount bought more than proportionately, because it permits a rearrangement of budgets in which the commodity in question as well as other commodities can share.

A commodity may have an inelastic demand at one range of prices and an elastic demand at another. Consider the demand schedule of Table 29.

Table 29 **DEMAND SCHEDULE OF VARYING ELASTICITY**

Possible prices of commodity	*Number of units purchased*
$5	2,000
4	2,100
3	2,500
2	4,000
1	9,000

This demand schedule is inelastic between $5 and $3, but it is elastic between $3 and $1. A comparison point by point would show that the degree of elasticity and inelasticity differs for each pair of prices in a demand schedule. In fact the most precise definition of elasticity is one which defines it at each price, so that conceivably a demand schedule could have as many elasticities as there are prices. In general, a commodity has an inelastic demand at very high prices and a less inelastic or an elastic demand at lower prices. The reason is that at very high prices only those with larger incomes are buyers; for them the price-income ratio is low, and their demand is thus inelastic. At lower prices buyers with smaller incomes enter the schedule and since there are then many more buyers, the demand becomes elastic.

Although it is broadly true that a demand is inelastic at high prices and elastic at low prices, there are instances in which demand becomes highly inelastic below some low price. Examples would be such commodities as shoelaces and aspirin, for which there is a relatively small amount that is needed or will be bought, even at lower prices. Advertisers often worry about the so-called saturation point in the sale of their wares; in part, the saturation point is the point at which the commodity has a great inelasticity of demand.

Income Elasticity

Still another relationship characterizes demand behavior: *income elasticity. Income elasticity is the relationship of proportionate changes in quantity bought—at given prices—and proportionate changes in buyers' incomes.* It has already been shown that a rise in income does not cause a uniform rise in demand for all commodities by an individual. Market demand schedules react similarly: if buyers' incomes increase, the demand for some commodities increases much more than others. Commodities differ in their income elasticities. One reason for a low income elasticity (i.e., a given increase in income causes a proportionately smaller increase in quantity bought) is that the price-income ratio is low: consumers have about as much of a commodity as they want and will not augment their purchases even with expanded incomes. Furthermore, many general classes of commodities can be ranked according to their "grade." Meat could be divided into grades (from the standpoint of the consumer) with tenderloin steak at the head of the scale and with plate beef or breast of lamb at the bottom. In wartime the problem of civilian meat supply was complicated by the heavy demand for the better grades of meat; the increase in the demand for these grades was much larger than the increase in incomes, i.e., the income elasticity of meat, especially the better grades, is high. In some instances, the demand for inferior qualities of a commodity actually falls off when income advances. Although shoes were rationed during the war, the government at times permitted the purchase of low-priced shoes without ration stamps, because the demand for low-priced shoes was not large owing to their low income elasticity.

The importance of the income elasticity of demand is best seen when it is related to the periodic expansions and contractions of national income and aggregate demand.[6] For purposes of examining income elasticity in this context, assume that the total consumer income increases by 10 per cent. The amounts of specific commodities bought might change as in Table 30.

Since the amounts of A and B increase more than 10 per cent the demand for them has a high income elasticity. Commodity C has an income elasticity of 1, i.e., the ratio of increase in income to increase in amount bought is 1 to 1. Commodities D and E have a low income elasticity, increasing only 5 and 2 per cent respectively when income rises 10 per cent. Consider now the amount of commodity F bought, which shows a decline

[6] See Chap. 6.

Table 30 **EFFECT OF INCREASED INCOME ON PURCHASES**

Commodity	Percentage increase or decrease in purchases
A	+25
B	+20
C	+10
D	+ 5
E	+ 2
F	− 5

of 5 per cent. Suppose that this represents an inferior grade of a commodity. As incomes increase some of the people who purchased this product before are now able to buy a better quality product, and demand, therefore, shifts to perhaps A or B. Commodities A and B are a larger proportion of national output than before, commodity C remains unchanged, and D, E, and F comprise smaller proportions. Relatively more labor and capital are of course devoted to the production of A and B than to the other commodities. On the other hand, when total income declines the demand for these products may be expected to decline unevenly. If the reduction in income is 10 per cent, the demand for A and B would fall by more than that, D and E by less than 10 per cent. The demand for F, however, might increase, because with lower incomes many consumers would shift their demand from a higher-priced commodity to F, the low-priced substitute. Income elasticities of demand, then, change the composition of the national output and the allocation of economic resources in a period of expansion or contraction.

Demand Interrelations

It was pointed out previously that a demand schedule for one commodity is valid for a given set of prices of other commodities. Exceedingly complex relations are involved between demand schedules. For purposes of simplification it is well to examine only two commodities which are substitutes for each other. The demand for one of the goods, A, is dependent on the price of the other, B.

Table 31 shows three possible prices of A, and below each price is the corresponding demand schedule for commodity B. The process could of course be reversed, and the demand for A made to depend upon the price of B. The demand for one commodity changes as the price of the other changes, assuming that all other prices (as well as incomes) are constant.

Table 31 **DEMAND SCHEDULES OF A COMMODITY IN TERMS OF PRICES OF A SUBSTITUTE**

A: $0.40 Demand for B		A: $0.35 Demand for B		A: $0.30 Demand for B	
Price	Am't in units	Price	Am't in units	Price	Am't in units
$0.60	1,000	$0.60	800	$0.60	600
.55	1,200	.55	1,000	.55	800
.50	1,500	.50	1,200	.50	1,000

Since the two commodities are substitutes, they are competitive with one another, and the relation of demand is therefore termed *substitutive.*

Another type of relation between two demands is termed *complementary.* Two articles are frequently associated in their use, e.g., phonograph records and needles, gasoline and oil. Such a demand relation may be shown hypothetically in Table 32.

Table 32 **DEMAND SCHEDULES FOR A COMMODITY IN TERMS OF PRICES OF A COMPLEMENTARY GOOD**

C: $5.00 Demand for D		C: $4.00 Demand for D		C: $3.00 Demand for D	
Price	Am't in units	Price	Am't in units	Price	Am't in units
$10	500	$10	600	$10	700
9	550	9	650	9	750
8	600	8	700	8	800

In this case the relation between the price of C and the demand for D is opposite to that for A and B in the substitutive relation. In the complementary relation, the lower the price of the one article, the more of it is bought and therefore the greater the demand for the other, because of the association of the two in consumption.

Neither the substitutive nor the complementary relation between the prices and the demands of two commodities is a 1 to 1 relation. Where the demands for two articles are associated, the proportions may be variable because of many factors. A tennis racquet and tennis balls have the complementary relation, but the racquet lasts longer than the balls and is more expensive. Oranges and grapefruit are examples of substitute, or competitive, commodities. If the price of one rises, the demand for the other increases, but not necessarily in the same proportion. Interdepend-

ences among demands extend beyond pairs of commodities. Whole groups of commodities are competitive with one another, and within such groups an internal competitive relation prevails. In a general sense, all commodities compete for the buyers' dollars, all are substitutes for one another, and thus all demand schedules are related. The complementary relation serves to group commodities into clusters, among which the competitive relation prevails.

Uses of the Demand Concept

The demand concepts are intensely practical. Businessmen need to know the demands for their products in order to estimate their needs for capital facilities (especially in new businesses), raw materials or finished commodity inventories, labor and services, and to gauge their production schedules and to set prices. To this end, many professional organizations engage in what is called market research to advise business firms which do not have research units of their own. Market research attempts by means of questionnaires, interviews, and other sources of information to determine the potential demand for new products and the future changes in demand for established products.

Calculations of the demand, with special emphasis on the elasticity of demand, were made for steel and automobiles before World War II. These studies were made primarily for the purpose of publicly justifying price policies in these industries. Calculations of estimated demand schedules are also made by businesses such as airlines, which must justify each new route to the government. In farm price policy the demand-schedule concept is vital. If the Department of Agriculture seeks to raise or to maintain the price of a farm product by reducing the amount sold, the questions arise, How much will the price rise and how much will the total income from the crop rise? Since the determinant is the demand schedule for the particular farm product, it follows that action to raise farm prices must be predicated on some notion of demand schedules. For agricultural commodities it happens that knowledge of the demand for farm crops is fairly extensive, for considerable research has been applied to this subject. Another government agency which is concerned with demand schedules is the Treasury, especially in its recommendations to Congress for new taxes. Suppose it recommends a rise in the federal tax on gasoline. How much more revenue will the rise in the tax bring in? Again, the answer depends on the nature of the demand for gasoline. So the Treasury, in planning its tax programs, must base them on estimates of demand schedules.

CHAPTER 14

Supply

Just as demand represents the behavior of the buyer, supply represents the behavior of the seller. Before examining supply and demand in their mutual determination of price, it is necessary first to analyze the relations of price and supply. The focal point of this analysis is the question of why commodities are offered for sale at different prices at different times and places, i.e., the question of relative scarcity. Wants and scarcities are the essence of the economic problem. Just as wants are basic to demand, so are scarcities to supply. Scarcity is that condition in which a commodity exists in quantities insufficient to satisfy all wants and is caused by particular scarcities of labor, materials, or equipment required to produce consumer goods.

THE SUPPLY SCHEDULE

A supply schedule is a means of analyzing the relations of prices to amounts offered for sale. *A supply schedule is a list of amounts of a commodity which would be offered for sale at various possible prices during a given time period.* The schedule in Table 33 may be taken as a simplified example.

Assume this supply schedule to apply to turkeys. It is shown that at 60 cents a pound a farmer will be willing to sell his entire flock of 50 turkeys, while at 20 cents, he will not be willing to sell any. He would sell different amounts at different prices, depending upon his expectations of future prices which are related in part to seasons of the year (Christmas and Thanksgiving, for example), and to other ways in which he can dispose of the turkeys (such as personal consumption). (The schedule under discussion pertains to a commodity already produced and later will be referred to as a "market-supply" schedule.)

Table 33 **A SUPPLY SCHEDULE**

Possible prices	*Number of units which would be offered for sale*
$0.60 (per lb.)	50
.50	45
.40	30
.30	10
.20	0

Supply, like demand, is always to be considered in relation to one or more prices, for the supply of a commodity cannot be measured or, in fact, be called existent, unless prices are specified. The farmer whose supply schedule is illustrated by Table 33 indisputably has a flock of 50 turkeys which can be counted in the field, but the number of turkeys in the farmer's field is a physical tally, not a supply in the economic sense. The supply is the number the farmer will make *available* to others according to price.

A supply schedule, furthermore, exists at a specific time, such as one day, or some period of time during which no significant change in supply influences take place. The supply schedule in Table 33 exists, perhaps on November 10, at a time when the farmer is wondering how many turkeys to sell in the Thanksgiving market. A week later or a month later, the schedule may change because the farmer's price expectations or his need for cash may have changed.

It should be noted that the relation between price and amount offered for sale in a supply schedule is direct: more would be offered at a high price, less at a lower price. This relation generally exists in supply schedules, although in some instances the relation of price and amount offered is inverse.

Increases and Decreases in Supply

By use of the concept of the supply schedule, the meaning of increases and decreases in supply can be examined. *An increase (or decrease) in supply means that at all possible prices more (or less) is offered for sale at each price.* Extending the illustration shown in Table 33, an increase and a decrease in supply is shown in Table 34.

The initial supply, schedule A, is derived from the prices and units of Table 33. The increase in supply (in schedule B) shows more units offered for sale at each price; it is assumed that the farmer has changed his

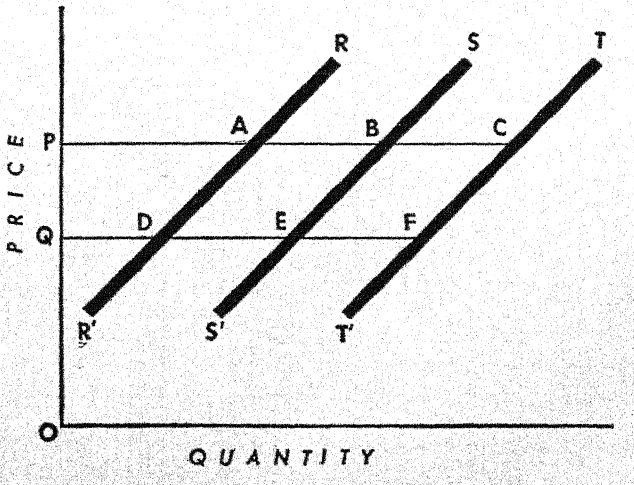
Table 34 CHANGES IN SUPPLY

Schedule A — Initial supply		Schedule B — Increase in supply		Schedule C — Decrease in supply	
Price per lb.	Am't in units	Price per lb.	Am't in units	Price per lb.	Am't in units
$0.50	45	$0.50	50	$0.50	30
.40	30	.40	45	.40	15
.30	10	.30	30	.30	5

opinion about future prices and since he now regards them less favorably is willing to sell more turkeys at each successive price. The decrease in supply (in schedule C) shows fewer units offered for sale at each price perhaps because the anticipated demand is high. It is of course essential to distinguish the differences between two supply schedules from the relation between price and amount offered for sale within a single supply schedule.

In Chart 10, three supply curves are shown. Supply curve *SS'* indicates that at price *OP*, for example, the sellers are willing to sell the amount *PB*, while at price *OQ* they will sell a smaller amount, namely *QE*. The curve *RR'* represents a decrease in supply: the amounts sold at both prices *OP* and *OQ* are smaller. Similarly the curve *TT'* represents an increase in supply. While the supply curves in Chart 10 are drawn as straight lines, they need not be. Supply curves can be of many shapes, depending on the circumstances involved.

Chart 10 INCREASE AND DECREASE IN SUPPLY



Elasticity of Supply

Elasticity of supply is the counterpart of elasticity of demand. *Elasticity of supply represents the relation between change in price and the corresponding change in the amount which would be offered for sale.* Consider the hypothetical examples in Table 35.

Table 35 ELASTICITIES OF SUPPLY

Schedule A *Very inelastic*		*Schedule B* *Inelastic*		*Schedule C* *Very elastic*		*Schedule D* *Elastic*	
Price	*Am't in units*	*Price*	*Am't in units*	*Price*	*Am't in units*	*Price*	*Am't in units*
$0.60	50	$0.60	50	$0.60	50	$0.60	50
.50	49	.50	45	.50	20	.50	40

Here are four supply schedules of varying elasticities. In both schedules A and B the decline in the amount which would be offered for sale at a lower price is smaller proportionately than the decrease in price, i.e., supply is inelastic. In schedule C, a decrease in price of 16 per cent is accompanied by a drop in amount offered for sale of 60 per cent while in schedule D the drop in amount offered for sale is 20 per cent for the 16 per cent decline in price. In schedules C and D the decrease in the amount offered for sale is larger, relatively, than the decline in price, i.e., supply is elastic.

Chart 11 ELASTICITY OF SUPPLY

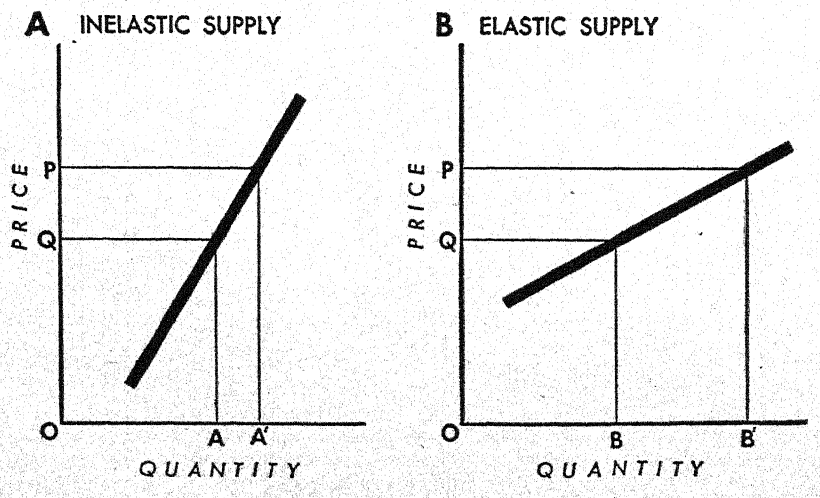


Chart 11 shows an inelastic and an elastic supply curve. It can be seen that a change in price, as from *OP* to *OQ*, causes a relatively smaller change in quantity in the case of inelastic supply, and a relatively larger change in quantity in the case of elastic supply.

There is, indeed, a condition of "perfect inelasticity." For example, suppose the farmer of this hypothetical case has only 50 turkeys and would sell all of them at 60 cents a pound. At any price above 60 cents he could not sell more turkeys because he has no more. The schedule at prices above 60 cents is perfectly inelastic because a higher price causes no change in the amount offered for sale.

COSTS OF THE FIRM

It is necessary in the analysis of supply to examine costs. Costs are a basic influence on supply and may indeed best be considered as obstacles to the flow of commodities: the higher the obstacle, the less is the flow. The terms *cost* and *costs*, however, are used so commonly that they require some definition. Costs cannot be limited to out-of-pocket expenses. Thus, a building contractor may say (to his intimate friends) that a house he has constructed cost him $6,000, meaning that he paid out so much for land, materials, and labor. But he may hasten to add that if depreciation on his equipment is counted in, plus something for his work in supervision, together with a fair profit, the actual cost is $9,000. This is an important point: the distinction between costs as out-of-pocket expenses and costs which include these expenses plus other items. Actually, it is not possible to lay down a precise definition of costs which suits all purposes. There are many purposes for which costs must be considered, and for each, costs can be looked at from a different point of view. What, for example, are the costs of traveling by automobile from one city to another? They might be stated in these different ways: (1) the money paid out for gasoline and oil; (2) the money equivalent of the gasoline and oil computed to be used on the trip, because this is conceivably different from (1); (3) either (1) or (2), to which is added a figure for depreciation on the car, insurance, taxes, and auto-club membership for the miles traveled on the trip; (4) the meals and other similar expenses incurred on the trip added to (1) or (2) or (3); (5) the difference in the price paid for meals, etc., over the sum which would have been expended had the man stayed home could be added to (1) or (2) or (3); (6) if the traveler thinks his time is worth something, he could add the value of the number of working hours lost on the trip to (5); (7) one of the foregoing figures could be compared with a figure valid for an

alternative means of transportation, say a bus ride, or sharing expenses with a friend, to arrive at a net cost for the auto trip. Possible methods of defining costs are not exhausted in the foregoing list. The principal purpose of the present analysis is the examination of costs as they affect supply.

The Short-Run Costs of a Firm

Costs of production are the prime influence on the willingness of a business firm to supply its product. Consider first a firm which has plant, equipment, machinery, and a labor force and which produces a commodity which it offers for sale. Its costs under these conditions are known as its "short-run costs," because *a period within which the equipment of a firm may be taken as fixed in amount is called the short run.* The long run is a period long enough for a firm to rebuild or expand its plant.

Suppose that the plant of firm A has been designed to produce, on a 40-hour week basis, 50,000 units per month, at an average cost of $7 per unit. This unit cost figure represents (1) the sum of the wages and salaries paid, (2) the amounts expended for materials, fuel, energy, professional services, etc., (3) depreciation charges, the monthly prorated cost of capital facilities used up in production, and (4) a return to the owners equivalent to what they could earn in the best alternative use of their capital. Some of these costs are independent of the actual volume of production; they would exist even if the plant produced nothing at all. Such costs are commonly termed the *fixed* or *overhead costs* of a business, and include the salaries of management, interest on borrowed capital, depreciation charges, the wages of permanent or skeleton staff employees (such as key foremen, janitors), and the costs of materials and services which continue regardless of the volume of production (e.g., minimum charge for electricity). The remaining costs constitute *variable costs*, which vary with the volume of production, and include the wages of the bulk of the employees and the payments for materials. These costs are shown for a hypothetical firm in Table 36.

In Table 36 different possible volumes of output for firm A are presented. Consider the relations between volume of output and costs:

1. As output is expanded the total of fixed costs remains constant. Fixed costs *per unit*, however, diminish as output increases. If output is comparatively small, say at 5,000 units, fixed costs per unit are relatively high at $20. But if the rate of production is advanced to 60,000 units a month, fixed costs per unit fall to as low as $1.66.

Table 36 COST-OUTPUT DATA FOR FIRM A

Units produced	Fixed costs	Total costs	Average total costs per unit	Average fixed costs per unit	Average variable costs per unit	Marginal costs
5,000	$100,000	$150,000	$30.00	$20.00	$10.00	—
10,000	100,000	170,000	17.00	10.00	7.00	$4.00
20,000	100,000	210,000	10.50	5.00	5.50	4.00
30,000	100,000	250,000	8.33	3.33	5.00	4.00
40,000	100,000	310,000	7.75	2.50	5.25	6.00
50,000	100,000	370,000	7.40	2.00	5.40	6.00
60,000	100,000	435,000	7.25	1.66	5.59	6.50
70,000	100,000	560,000	8.00	1.43	6.57	12.50

2. Total costs increase as output is expanded. This is shown in the third column of Table 36. But total costs do not, except in rare cases, increase at the same rate as output. For firm A, *average* total costs per unit diminish up to an output of 60,000 units per month; but for the output of 70,000 units, they become higher. The explanation of this is to be sought in the behavior of variable costs as output increases.
3. Where output is small, average variable costs per unit are comparatively high, because too little labor is employed relative to the capital facilities used. Production is not "efficient" inasmuch as the firm cannot take full advantage of the possibility of specialization among its employees. When output is expanded, however, average variable costs per unit may diminish. In the example of firm A given in Table 36, average variable costs per unit decline at first, then increase as output becomes larger. This relation between average variable costs per unit and output is probably typical, but does not hold true for all firms; in some cases these costs may be constant for some range of output before finally beginning to rise. For large volumes of output—greater than the volume of output for which the plant was designed and for which average *total* costs are at a minimum—average variable costs per unit rise with output because there is too much labor in proportion to plant, because perhaps overtime wages must be paid, because of crowding, increased fatigue, and because of higher rates of breakage or spoilage.[1]

[1] For a fuller discussion of why variable cost per unit changes, it is necessary to consider the combination of the factors and the law of diminishing returns. A discussion of these subjects is given in Chap. 17, and pp. 395-402 may be read at this point if the reader wishes to pursue this matter further.

4. Marginal cost is the increase in total cost consequent upon the production of an additional unit. Thus, in the case of firm A, total costs rise from $370,000 to $435,000 when output is expanded from 50,000 to 60,000 units per month. This is an increase in total costs of $65,000 for an additional 10,000 units; thus the marginal cost is $6.50 per unit. Strictly speaking, the marginal cost so computed is an *average* marginal cost. Marginal cost proper is the cost of one more unit. In the case of firm A, marginal cost is constant for part of output, and then rises as output is increased. Marginal costs need not always exhibit the behavior shown in Table 36; although they usually decline to a certain point and then rise, the gradations may be small throughout the range of output rather than as large as they are in the example.
5. Average total costs per unit are the full costs of production per unit and are the sum of fixed costs per unit and of average variable costs per unit. For firm A, average total costs per unit decrease as output expands, primarily because of the decline in fixed costs per unit. For high levels of output, however, average total costs per unit rise, because the increase in average variable costs per unit more than offsets the continuous decline in fixed costs per unit. The rate at which average variable costs per unit decline depends on the proportion of fixed costs to total costs, and upon whether and at what rate average variable costs per unit decline.

The distinctions between the various types of costs and the way in which they vary with output for any firm have an important bearing on the amounts which a firm will produce at various possible prices in the short run. How do firm A's costs affect the amount of its product which will be offered for sale? This question can be answered by examining the relations between the firm's costs and the possible prices it could receive for its product.

Let it be assumed that firm A can sell any amount of its product at any price which might prevail. Suppose that a price of $5 per unit were to exist. At this price firm A would face losses because the price is below its average total costs per unit for any range of output. However, if firm A did produce 30,000 units per month and sell them at $5 each, its loss would be no greater than its total fixed costs; the price of $5 covers the average variable cost per unit, which is also $5 for an output of 30,000 units. Thus, firm A's loss would be no greater than if it did not produce at all. Therefore, at any price above the lowest average variable cost, the firm will gain more or lose less by producing than by not producing. Suppose now that

the price rises to $6.50. At this price firm A would find it worth while to produce 60,000 units, because although it suffers a loss—$45,000 per month—the loss is smaller than the $100,000 monthly loss which would exist if the plant were to be shut down altogether. If the price rises to $7.25, firm A will be in a position where it no longer has a loss, because $7.25 is its average total cost per unit for an output of 60,000 units. However, before the firm will expand output to 70,000 units, the price must be at least $12.50.

The figures in the last column of Table 36 show marginal costs. For the output of 70,000 units a month marginal costs are $12.50; and while firm A would make a profit at a price of $10 per unit, it would not pay to expand to an output of 70,000 units a month, because the addition to its total revenue, namely $10 per unit, is less than the unit addition to its costs which is $12.50.

To sum up the position of firm A:

At a price of $5.00 it will produce 30,000 units
At a price of $6.50 it will produce 60,000 units
At a price of $12.50 it will produce 70,000 units

At prices in between those cited in the example, the firm will produce amounts determined by comparing additions to cost (marginal cost) with additions to revenue (in this case, prices). Amounts which will be produced at each possible price are therefore determined by finding what amounts have a marginal cost equal to the possible price under consideration. Therefore, the supply schedule of the individual firm which takes the price to be given is the marginal cost schedule above the point of minimum average variable cost. (No output would be supplied at prices below minimum average variable cost.)

The costs of firm A and its price-cost relationships have been described. A more generalized description of price-cost relationships may now be undertaken with the aid of Chart 12, showing the various unit costs already described. This chart is not derived from the cost figures of firm A because it shows fine gradations in cost and output which would require too much space if illustrated in table form; however, it exhibits the general relationships already described for firm A.

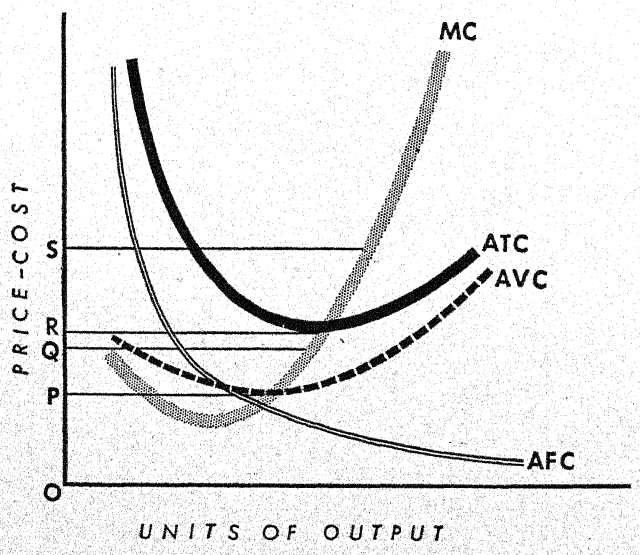
In the chart, *OX* measures units of output produced in a given period of time and *OY* measures costs per unit of output. Costs are all premised upon given factor prices, and factor prices are the same for all levels of output. The vertical axis may also measure selling prices and for present purposes several prices (*P, Q, R, S*) are shown. The four cost curves shown in Chart 12 are *costs per unit of output* and are as follows: (1) *AFC* is the average

fixed cost per unit, (2) *AVC* is the average variable cost per unit, (3) *ATC* is average total cost per unit, and (4) *MC* is marginal cost.

As output increases the average fixed cost declines because the total of fixed costs is distributed over, or allocated to, increased units of output. Thus, *AFC* declines sharply at first and continues to fall as the number of units becomes larger. Average variable costs, shown as *AVC*, decline over part of the range of output and then increase, for reasons already given.[2]

Average total costs per unit—the sum of *AFC* and *AVC*—at first decline and then rise, as the increase in variable costs per unit offsets the decline in average fixed costs per unit. *MC*, measuring marginal cost, is of utmost importance; it is the addition to total costs arising from the production of an increment of output. It will be noted that *MC* crosses *AVC* and *ATC* at their lowest point. This relationship arises because these average costs must increase when marginal cost exceeds the average. Thus at the point of intersection of *ATC* and *MC* both costs are equal; if beyond this point the cost (i.e., the marginal cost) of an increment of supply increases, the average cost of this increment plus the previous amount of output is necessarily increased.

Chart 12 PRICE-COST RELATIONSHIPS OF A FIRM



[2] See also footnote 1, p. 319.

Consider now the relationship between unit costs and market price. Below price *OP* no production would be undertaken, for at lower prices the producer would not cover his variable costs. At price *OP* and above the commodity would be produced because variable costs at least would be covered by the price. Suppose the price is *OQ;* the firm would then produce up to the point at which marginal cost equals price. The horizontal line from *Q* to *MC* measures the amount produced. No more than this would be produced for beyond this point the cost of the additional output (marginal cost) would exceed the selling price. No less than this would be produced either, because at any lower output more could be added to revenue than to cost by increasing output.

Price *OR* equals the minimum average total cost, and the horizontal line drawn from *OR* measures the amount produced. No more than this would be produced at the price *OR,* again because the cost of producing each added unit beyond this point exceeds the selling price. If the price should be *OS,* however, output would be expanded as indicated. This graphic illustration thus makes clear that output will be produced (1) where price equals at least the minimum average variable cost, (2) more than this will be produced where price is above this point, and (3) at any higher price output will be limited to the amount at which marginal cost and price are equal.

The Planning Cost Schedule of a Firm

Firm A is a plant with a given amount of facilities, and its unit costs of production were based on this "fixed" amount (and condition) of facilities. But costs vary with productive capacity or size. Therefore in planning the size of a firm it is necessary to consider the relationship between size and average unit cost. If the owners of firm A were to start afresh and consider the size plant they should now build, they would endeavor to plan its size in order to obtain the lowest average total unit cost. They would be faced with alternatives like those shown in Table 37. For varying sizes of plant the minimum average total cost per unit ranges from \$8.80 to \$5.50. A *planning cost schedule* such as this would enable the owners of firm A to determine the most suitable size.[3]

[3] The consideration of costs as related to size of plant, together with the producer's expectations of receipts from the sale of output, govern the volume of investment. See Chap. 5, pp. 109-19, for a discussion of this point as it bears on the level of national output and income.

Table 37 **COSTS IN RELATION TO SIZE OF FIRM**

Size of firm A	Monthly output in units for which designed	Total costs per month	Minimum average costs per unit	Marginal costs per unit
1A	25,000	$220,000	$8.80	—
2A	50,000	385,000	7.70	$6.60
3A	100,000	550,000	5.50	3.30
4A	150,000	990,000	6.60	8.80

In these figures, the lowest average total unit costs decreases as the size of the firm increases—but only up to a point, for in 3A minimum unit costs are lower than in 4A, the largest size firm. This pattern of decrease to a minimum point, followed by an increase, is generally characteristic of single enterprises. If costs per unit were the same regardless of the size of the enterprise, no minimum—no efficient size—could be determined. On the other hand, if costs declined continuously as size expanded, costs would equal zero if the size of the firm were increased enough. Likewise, if costs increased from the beginning, an infinitesimally small firm would be recognized as the most efficient. How, then, is the minimum point determined? In some enterprises, most branches of agriculture for instance, the minimum cost is reached when size is small. In other enterprises, steel production for instance, minimum cost is reached when size is comparatively large. There is a more or less definite minimum-cost size for each industry, at any point of time. When, over a period of time, technology changes, the minimum-cost size of a firm may become larger or smaller. The development of the assembly-line technique in automobile production, for example, made large assembly plants more economical than small ones. However, farm tractors developed during the 1930's enabled the family-sized farm, in many areas, to maintain its competitive position against large-scale farming.

The basic reason for the common pattern of decrease-minimum point-increase in unit cost as determined by plant size is the fact that management, or "top-management," is indivisible. As a firm expands in size its land, labor, and capital are increased, but management—the entrepreneur—cannot be increased. Ultimate authority and responsibility cannot be divided. In other words, minimum unit cost is achieved by the most effective allocation of resources; and past a point, management cannot be allocated most effectively. Allocation of resources is, as was previously stated, the main determinant of production. If less land, labor, and capital are used than the "ideal" amounts, unit costs must be higher; if more are used, costs are also higher. There are great differences, from industry to industry, in

the way in which costs decline to their minimum point and rise again as size is increased. In retailing, for example, there is little difference in costs between large and small units. A mail-order house operates at lower costs than a small department store, but the cost advantage is apparently not very great. On the other hand, in electric power production, the differences in unit cost between large and small plants are very considerable.

In Table 37, which shows possible sizes of firm A, the costs include a minimum profit, which is a "profit" just large enough to induce investment in the size plant indicated. Size 2A is the size of plant used in Table 36, which showed fixed and variable costs. In Table 36, total costs for 50,000 units were $370,000; in Table 37, total costs are given as $385,000, or slightly higher. This increase is arbitrarily selected here as a minimum profit. It was not included earlier, because *ex hypothesi* the plant already existed; the minimum profit is the minimum *inducement* to maintain production over long periods of time. If a plant already exists there is no need to consider the inducement to construct it. Minimum profits, however, are to be included in costs when it is a question of deciding whether plants are to be built, i.e., whether investment is to be undertaken.[4]

Marginal costs, given in the last column of Table 37, show that while the minimum average cost for size 4A is higher than for size 3A, the marginal cost is much higher. This focuses attention on the extra cost per unit for size 4A. If the businessman were to plan on increasing the size of the firm from 3A to 4A, he would be foolish to look only to his average costs. Suppose that he was getting a price of $5.50 for his product while his plant was of size 3A. If he expected the price to rise to $7 and therefore increased his size to 4A because average costs at 4A would be $6.60, he would actually lose money. The reason is that the $7 price would not cover the extra or marginal costs of the extra or marginal output of the larger plant. Actually, he would have to expect a price of at least $8.80 before he would expand his plant from size 3A to size 4A. The expression "marginal cost" is not commonly used in the business world, but its concept determines business decisions nonetheless. The extra or additional cost of doing something is its marginal cost. Suppose the price of milk received by a dairy farmer rises. He knows he can get more milk out of his herd by feeding the cows more high-protein fodder. Will it pay to get more production now that the price of milk is higher? The dairy farmer can calculate the additional cost of additional milk, and compare this with the additional income it will bring in. A calculation of this kind is one in which marginal cost is compared with price, even though the farmer knows nothing of the term marginal cost.

[4] See Chap. 21 for an extended discussion of profits.

THE RELATIONS BETWEEN COSTS AND SUPPLY SCHEDULES

Thus far two principal concepts have been developed in the analysis of supply: the supply schedule for a commodity and the supply schedule of a firm. Both of these concepts are applicable only under conditions in which individual sellers themselves do not determine the prices they receive. The supply schedule for a commodity must be related to costs. There are a number of complex relations between supply and costs, and it is easiest to examine these relations by showing how they differ over periods of time. To answer the question, "How does the supply—the amount of a good offered for sale—depend on cost?" one must consider whether the supply is actually now in existence (market period), or whether goods have yet to be produced with plant and equipment now available (short run), or whether the plant and equipment have first to be produced so that goods can be produced (long run). The supply-cost relation will be discussed in terms of (1) the immediate market period, (2) the short run, and (3) the long run.

The Immediate Market Period

The immediate market period is any single period in which a seller or sellers have on hand a stock of goods ready for sale. The period might be of a few minutes' length or a few months' or longer. The production of goods for future sale is not involved in the market period, for it relates only to commodities which have been produced in the past. Therefore, it is evident that *any period in which quantity to be sold is taken as fixed (on hand) is by definition a market period.* A seller in the market period might be a dealer in wheat or cotton, a farmer who has already harvested his crop, a manufacturer who has inventory of his manufactured products, a retailer with goods on his shelves.

The price at which the owner of a commodity is willing to sell in a market period is not determined by the production costs of that commodity. The seller's supply schedule for the market period is the same as the supply schedule shown in Table 33. In some cases costs may influence supply in the market period if the seller stands obstinately on the ground that his wares cost him so much and that therefore he will not sell for less, i.e., he will not sell below the amount of his production costs. Such an attitude ignores the fact that the costs of anything that exists are *past* while its price is *immediate* or *anticipated.* And the prices at which a seller will dispose of goods he has logically depend on his estimates of the prices he expects will

exist. Nevertheless, the attitude that prices "should" bear some definite relation to costs is very common and deeply rooted. This attitude is a fact, illogically conceived or not, and must be taken into account as such.

The Short Run

Consider now the relation between costs and supply in the short run. *The short run is a period long enough to make the amounts supplied depend upon production but short enough to preclude any change in the size or number of plants.* The actual length of time in the short run may be as short as a few weeks or as long as a few years, depending upon the nature of the commodity or service. In general, the short run is lengthier in duration for those commodities whose production requires a heavy investment in specialized capital equipment than for those products or services whose production requires only equipment which can be readily transferred with little or no loss to the production of other products.

The analysis of the costs of a firm can be brought to bear directly on the short-run supply schedule of an individual seller. In the first place, the prices in his supply schedule are expected, or anticipated prices: some time must always elapse between production and sale. The only exception—and it is an important one—applies to those products and services whose production and consumption occur simultaneously, such as electrical energy. But even in service industries *anticipated prices* govern the volume of production and hence supply.[5]

The short-run supply schedule depends upon the marginal costs of the firms producing the commodity. It has been shown that at any given price the firm will produce an amount for which the addition to total cost (marginal cost) is equal to the addition to total revenue (price). The individual firm's supply in the short run becomes perfectly inelastic at a price which calls for the entire output of which a firm is physically capable. At a very low price, a firm will not produce at all; the supply is zero. For supply to become zero, the expected price is *below the lowest level of average variable costs.* Naturally, the owner of an establishment loses money (the amount of his total fixed costs) when he shuts his plant down completely, but if the price is so low as not to cover variable costs, he would lose not only a sum equal to his fixed costs but, in addition, the difference between total variable costs and the total revenue from sales. If price were high enough to cover variable costs, but not so high as to cover

[5] The importance of short-term price and profit expectations on production and the level of national output and income was discussed in Chap. 5.

total costs, the loss from producing at such a price is less than the loss from not producing at all. The *firm's* supply schedule, therefore, shows that no output would be produced at a price below minimum average variable cost, that above this price the firm will produce outputs which equate marginal cost and price—the higher the price the larger the output—but that prices above some level will not call forth more output because the firm cannot produce more. The short-run schedule for an *industry*—a number of firms producing a standardized commodity—is derived from the short-run supply schedules of the firms. It is the sum of the amounts which each firm would produce at each possible price.

Nevertheless, a firm may be unwilling to supply its product at any price less than its total costs, especially if it expects a low price to continue for only a short while. The policies of business firms vary widely and even the same firm may have different policies at different times. Thus, a firm may at one time be very eager to get a large order and may offer to supply the goods at a price below its full costs, but at another time it may decline to sell anything at below costs. It is fairly common for businesses to add some figure like 10 per cent to their total costs and consider this the minimum price to be accepted.

The Long Run

The long run is a period which permits producers in an industry to adjust the sizes of their plants and which allows the number of producers in the industry to increase or decrease. The long run must be considered more as a set of *conditions* than as a period of chronological time. These conditions are similar to those confronting a businessman when he is making plans to produce a product which involve decisions as to kinds and amounts of capital facilities, type of organization, and size and composition of labor force, assuming certain factor prices and a given technology. Under these circumstances his decisions will involve only the adjustment of productive facilities to possible prices for his product. Each possible price is expected to prevail indefinitely. The conditions assumed here contrast sharply with those described in Chapter 5 (page 116) where the businessman was shown as confronting uncertainties over a period of chronological time in which factor prices, technology, and product prices are all subject to change. The long run, therefore, is an abstraction which sets aside all other influences except the *relation* of prices of the product to the kind and size of productive facilities.

Cost changes found to occur in practice can be grouped into three types:

(1) changes in the prices of the factors of production; (2) changes resulting from variation in the proportions in which the factors of production are combined; and (3) cost changes due to alterations in technology. The third type is particularly important. New machines and devices and new processes and methods have lowered costs of many products very sharply. The effects on cost of changing *technology* are often confused with the effect on cost of the second type, a *different scale of production.* In fact, either or both of these effects are commonly confused with the effect on costs of increased *capacity utilization.* Decreasing unit costs with increasing production are, therefore, caused by three main factors, separately or in combination: (1) a more economic proportion of factors of production, as when production is expanded from 50 per cent of capacity to 80 per cent of capacity; (2) a larger scale of operation, as when a bigger plant is built which takes advantage of economies not possible in a smaller plant; (3) use of improved technology over a period of time. The first reason is applicable to the short run; the second to the long run; and the third, although of great practical importance, is for purposes of simplicity abstracted from the analysis of the long run.

To illustrate, suppose the total unit costs of production of an article are $2 when the plant producing the article is operated at 50 per cent capacity. When the plant is run at capacity, costs per units are $1.50. This is what people mostly have in mind, although incorrectly, when they talk about decreasing costs. But when the item is produced in a larger plant designed to take advantage of a given technology and if the plant is run at capacity, costs may drop to $1. This is an example of long-run decreasing costs. If research and ingenuity subsequently improve production processes, the cost may be brought down to $0.50. In practice, the influences of the long run and of changes in technology are intermingled with those of short-run variations in output; techniques improve while larger plants are being built, and in any one period a plant may operate at 70 per cent or 90 per cent of its capacity.

SUPPLY SCHEDULES FOR GROUPS OF SELLERS

The supply schedule of an *individual seller* has been treated thus far. For some purposes of analysis, it is desirable to group together individuals' supply schedules into (immediate) *market supply schedules. A market supply schedule is the sum of the amounts which all sellers of a particular commodity would be willing to offer at a particular time at various possible prices.* The commodities offered for sale must be identical; otherwise, a

simple error in measurement is committed. Thus, for a commodity such as a certain grade of wheat, a market supply schedule for the Chicago market on a given day could be established, if the minds of the sellers of wheat were fully explored. It is of course usual to talk about the (market) supply of wheat even though wheat comes in several grades, as do cotton or beef. The supply on the market of the staple agricultural commodities may change from hour to hour. The physical amount on hand obviously cannot change much from hour to hour, but the market supply schedule—the willingness of the sellers—changes drastically. Into the markets for agricultural products there pour news and rumors constantly: crop forecasts and reports, foreign developments, announcements of new government policies, and other information which cause sellers to revise their positions. For perishable products, the supply schedules of sellers change as the product approaches a decayed condition. For example, in a vegetable and fruit market place, supply schedules for perishables increase sharply during Saturday evening if the sellers have no storage facilities.

The *short-run supply schedule for a group of sellers* can also best be illustrated by reference to an agricultural product. Take some one crop. Before the planting season farmers must make a guess at what next year's price will be. If they believe it will be high, they will plant a large acreage of the crop in question. If they think the price will be low, they will collectively plant a small acreage and devote other acres to a crop whose estimated future price is not relatively so low. Thus the short-run supply schedule consists of the estimated future prices with amounts to be planted associated with each price. When the short-run supply schedules of many different producers are compared, it is to be expected that individual supply schedules will differ from each other because of varying costs of each producer. *In a short-run supply schedule each price shows the total amount of the commodity which would be produced at that price.* A short-run supply is illustrated in Table 38.

Table 38 **SHORT-RUN SUPPLY SCHEDULE**

1. Expected price	2. Total planned production in units	3. Amount planned to be produced in units by firms		
		A	B	C
$1.50	50,000	25,000	15,000	10,000
1.25	35,000	20,000	10,000	5,000
1.00	20,000	15,000	5,000	—
.75	10,000	10,000	—	—

Columns 1 and 2 show the supply schedule for all sellers combined, and since it is the short-run situation which is illustrated here, column 2 is headed by "total planned production." Column 2 is the total of A, B, and C in column 3, which are the supply schedules of three producing firms (although there may be 300 or 3,000 or more). At $0.75, A is willing to supply a small amount; at $1, A is willing to supply more, and B is willing to supply a small amount; at $1.25 A is willing to increase production still more, B is willing to supply more, and C will now supply a small amount, and so on. Hence, as the price becomes higher, additional producers enter the market and the amount planned to be produced is successively increased.

The *long-run supply schedule of a commodity shows the amounts which would be produced when both the number of producers and the sizes of their plants are adjusted so that minimum average costs are equal to each possible price.* The long-run supply schedule may be conceived in terms of the growth of an industry. If an industry expands its output over a very long period assuming that the supply of factors and technology do not change, will its product be available to buyers at a higher, a lower, or the same price? It is traditional in economics to classify industries from the standpoint of their long-run supply schedules as "increasing-cost," "decreasing-cost," and "constant-cost" industries. As total output is expanded (the sum of the output of all firms), the minimum average total cost rises, remains constant, or falls, for industries of increasing-, constant-, or decreasing-cost industries. However, much doubt has been cast on the logic and usefulness of this traditional classification. It is now generally accepted that all industries are increasing-cost industries, i.e., a whole industry can expand its output only in face of higher unit costs in the long run, assuming, of course, there are no technological changes. The reason is that an expanding industry draws resources away from other industries and thus must pay higher prices for the resources and raise its own costs. In any case, it is the practice in modern economics to confine the concept of a supply schedule for an *industry* largely to agricultural commodities, mainly because farm products are standardized and therefore the many sellers can be grouped together in the supply schedule of an industry. In the analysis of farm products there has been little or no disagreement that they are produced in increasing-cost "industries."

Naturally, between industries there are different rates at which costs increase in proportion to long-run expansion. For industries where there is relatively little capital equipment and much labor, such as clothing manufacture, the increase is slight. Other industries, like lumber production, which depend on scarce resources available to the industry in inelastic supply, often show sharp increases in cost.

The concept of increasing-cost industries stands forth clearly only if technological change is ignored. If technological progress were to cease, then those industries which expanded would do so (assuming full employment in the long run) only at the expense of others, and hence their unit costs would rise. But technological improvements occur, and although they do not "disprove" long-run increasing costs, they obscure the tendency of costs to increase in the long run. Thus, for example, over the period of a generation the price at which a given amount of a mineral is supplied may decline, because of technological improvements in production; but at the end of the generation, the best mines may be worked out and then the mineral must be extracted from poor mines. Had there been no technological improvements the costs of production of the mineral would have risen. These two forces—technological change and long-run increasing costs—are always opposed, but thus far in the industrial history of the nineteenth and twentieth centuries technology has been the dominant force. Economists a century ago believed that increasing costs would prevail over the decline in costs resulting from technology: their pessimism gave economics the name "the dismal science."

Interrelations of Supply

Supply schedules, like demand schedules, are interrelated. To begin with a simple case, consider what is known as "joint supply." *Joint supply means that two commodities are produced together from the same natural or technological process:* wheat and straw, beef and hides, wool and mutton, gas and oil, lumber and sawdust. Thus all saleable by-products are results of joint supply. Since the joint production of the two commodities is unavoidable, their supplies are inseparably connected. If the market supply of beef increases, so must the supply of hides; if the supply schedule for beef alters, so must the supply schedule for hides.

The *elasticities* of supply for joint products may, however, differ sharply. A sawmill, for example, produces and sells lumber and sawdust. Assuming for present purposes that all grades and types of lumber may be considered as a single commodity, it is obvious that each lot of board feet of lumber produced will cause a number of tons of sawdust to be produced. The short-run supply schedule for sawdust is then very inelastic, because the decision to produce sawdust does not depend on the expected price of sawdust but rather on the output of lumber (which depends on anticipated prices). Since sawdust is used mainly as a fuel and lumber as a building material, the actual and anticipated prices of these two products are not

directly related; consequently, if there is a rise in the expected price of lumber more sawdust will result, even if there is no rise in the price of sawdust. In short, then, the supply of sawdust must be inelastic—a change of its price has little effect on the supply—because of its dependence on the output of lumber. Conceivably, technology could make sawdust the more valuable product. In this case lumber would be the by-product, and the foregoing analysis would be reversed.

Two commodities may be supplied under the conditions of "rival supply." *Rival supply means that less of one commodity is produced when more of another is produced,* because the two commodities are rivals for the same resources. Aviation gasoline and ordinary motor fuel are found in a rival supply condition. If a business corporation has facilities to produce both commodities, in the short run, when the production facilities are fixed, the supplies of the two commodities depend on their prices. Even when both prices are high, if aviation gasoline is higher than motor fuel it will be produced primarily, to the subordination of motor fuel even though the latter's price is high when considered by itself. In a sense, all commodities—with the exception of cases of joint supply—are rivals, i.e., each commodity uses resources that could be used in producing another. This condition, which is obscured in periods of depressed employment, existed during World War II, and is actually always present. For example, there is a limited amount of arable land, and hence all crops and animal products compete for available land.

The proposition that supply depends on actual or prospective prices has been developed in its important aspects up to this point. There are, however, instances in which it is not true that for each actual or possible future price a different supply is forthcoming. If an industry's long-run costs were constant, the relation between supply and price would be one in which a large amount would be forthcoming at the same price as a small amount. But this is a special case. However, the same relation—different amounts supplied at some one price—holds for individual firms in some instances. Some products sell for many years at the same retail price: magazines, soft drinks, cosmetics, in fact most branded goods, customarily sell at consistent prices from year to year and usually from one city to another. Sellers of such commodities have found it advantageous to establish firm prices which they may adhere to even through wars and depressions. However, in these cases the sellers have some control over their own prices; the prices are seller-determined rather than market-determined prices. The concept of a supply schedule does not apply in such cases.

Aggregate and Particular Supply

The analysis of the supply of a particular commodity has been described up to this point; consider now the relation of particular supply to changes in national output, income, and employment.

When national output expands it is not to be expected that each type of commodity and service expands in proportion: the output of some industries increases relatively more, others relatively less, than total output; still others may remain stable, and some may actually decline in the face of a general expansion. It has been pointed out that this diversity in rates of change may result because as total income rises there are marked differences in the income-elasticities of demand. Moreover, each industry is faced with conditions peculiar to it. Given their existing capital facilities, expansion of output of some industries encounters sharply rising costs; in other industries costs may change little if at all; in still others, unit costs of production might decline as volume increases. Some industries are able to expand their capital facilities—and therefore their output—over the course of a few months; others require several years. Expansion brings in some industries such "bottlenecks" as skilled-labor shortages and lack of sufficient materials, while others encounter these difficulties much later. Eventually, as a condition of relatively full employment is reached, the supplies of most products become highly inelastic, for the economic resources essential to increased production are not available.

Likewise, as national output, income, and employment decline, the effects on the production of specific commodities and services are dissimilar. The supply of some things, e.g., capital equipment and durable consumer goods, falls off sharply; other commodities and services decline only a little, and some increase. Night clubs may lose much of their business because their former patrons are led by reduced incomes to go to the movies, or entertain at home. As noted in the preceding chapter, a general decline in consumer spending is unevenly distributed; changes in supply are also uneven. Moreover, some industries are able to contract supply quickly, while in others supply is relatively inelastic.

SUMMARY

The supply schedule is an analytical device to represent the relations between prices and amounts which will be sold. In the market period, supply depends primarily on price expectations. In the short run, supply depends

on the relation between expected prices and short-run costs. In the long run, supply depends on the relation between expected prices and full costs which include a minimum profit.

CHAPTER 15

Prices and Markets: I

In the two preceding chapters demand and supply were examined separately; in this chapter demand and supply are viewed together as determinants of price. In the preceding analysis the level of a particular price was necessarily taken for granted; and though the effects of prices on amounts which would be bought and amounts which would be offered for sale were examined, the causes of price changes were not discussed. The causes of price differences and price changes form the subject matter of this and the next chapter.

Prices are determined in markets. *A market is any area in which buyers and sellers are in touch with one another and therefore in which the forces of demand and supply affect prices.* A market is not necessarily a physical place, like an auction room, or the "floor" of an organized exchange. Some markets have no special physical location at all; buyers and sellers simply telephone one another their offers to buy and sell at stated prices. A market may be limited geographically, as in a country store, but it may also be international in scope, as in the case of certain securities. Furthermore, markets can be analyzed from different points of view. For example, "the retail market" in different levels of analysis may refer to (1) the total market for consumer goods in the nation, (2) the national sales potentialities for a single commodity, or (3) the retail stores in a city which may be treated as one unit, or divided into commodity groups, areas, types of stores, incomes of the buyers, or in other ways.

In every market there is competition among the buyers, or among the sellers, or both, excepting only such instances as when there is only one buyer and only one seller of a commodity. The next thing to probe into, then, is competition. "Competition" is another word in common usage whose

meaning is usually taken to be self-evident. However, consider these expressions: unfair competition, free competition, cut-throat competition, wasteful competition. The modifiers in this list (which could easily be expanded) suggest, at the very least, that competition is a many-sided institution. It is well, therefore, to be explicit in distinguishing what form or aspect of competition is being referred to in a particular economic analysis.

Competition is varied in form: it may take the form of price competition or of quality competition. Competition may exist among a few sellers or buyers, or among hundreds of thousands. In a general context, *competition is rivalry in buying and selling goods and services in which the actions of each buyer or seller are affected by the actions of rival buyers or sellers.*[1] Competition in this sense is a vital part of a price economy. Together with the institutions of private property, representative government, and freedom of enterprise, competition determines the entire character of economic activity in such economies as those of Canada and the United States. Competition is literally the organizing principle of the economy. In this sense, "competition" and "the market" are interchangeable with "price system" as a definition of the economy.

Competition is a complicated institution because it exists in a great many different forms. Suppose a businessman wants to get ahead of his competitors, to attract more customers to himself. He can do it in many ways: lower his price, spend more on advertising, improve the quality of his product, give better service or quicker delivery or easier credit terms, and so on, ad infinitum. There are many ways to compete. But the businessman who seeks to best his rivals must always know exactly who his competitors are. A different kind of example is the competitive position of a farmer who grows a staple crop such as wheat, corn, or cotton. The farmer will not seek to improve his position by lowering his price, because he does not set the price himself and because he can sell all he wants to sell at the prevailing price; nor will he spend more on advertising because neither he nor any other farmer has any advertising expenses at all; nor will he engage in most of the other competitive practices of retailers or manufacturers. The farmer is engaged in competition but his means and methods are radically apart from the businessman's.

It was said that the actions of buyer or seller are limited by his competitors, i.e., other buyers or sellers. Thus one aspect of competition is the *number* of rivals for any buyer or seller. It is of great concern to a seller whether he has six competitors, or sixty, or six hundred thousand. If the

[1] Cf. J. M. Clark, "Toward a Concept of Workable Competition," *American Economic Review*, Vol. xxx, No. 2, Pt. I, June, 1940, p. 243.

seller has six competitors, he watches all six closely, knows the traits of each one fairly well; if he has sixty, he will probably closely observe only a few of them, although for some purposes he will have to take into account the general tendencies of "the industry"; when he has six hundred thousand, he does not even think of them as "competitors," but as "the market," a great mass in which individuals' identities are indistinguishable. If one competitor is limited in his actions by what his rivals do or might do, it is necessary also to take into account the *substitutability* of one competitor for another. If a manufacturer has six or sixty competitors, does he face the same degree of rivalry from each one, or are some of them close competitors and others less close? Actually, competition is usually not uniform among all competitors; it is typical for instance that competitors in the same general field tend to align themselves in small groups. A city may have six department stores; one caters to the "carriage trade," selling high-priced articles only; two sell inexpensive grades of merchandise only; and the three remaining stores carry some high-priced, and some low-priced, but mostly middle-priced articles. All six of these department stores are in the same business, all compete with one another, since some shoppers may go from one to another to search for goods; yet, the two stores selling cheaper goods are direct and close rivals, as are the three appealing to customers wanting medium-priced goods, and one tries to maintain a dignified aloofness.

The substitutability of one competitor for another is also a matter of the relative capacities of competitive businesses. In many branches of manufacturing the corporations composing an industry are of quite unequal size; a big corporation competing in the same field as a small one has many competitive advantages over its smaller rivals. Because of the size factor, among others, *direct* rivalry between two business firms does not imply matched strength or power between them.

MARKET SITUATIONS

In order to clarify the analysis of the many types of market situations, it is convenient to classify them as follows:

1. *Pure competition.* This is the term to apply to those competitive markets characterized by a large number of sellers or buyers.[2] A *large* number of sellers is a number of such magnitude that no one seller has

[2] The term *perfect competition* is sometimes used in the same sense as pure competition. In this book the word "perfect" is not employed because it is reserved for those cases in which there is perfect knowledge of the market and perfect mobility of the factors of production.

any appreciable influence on price. Under conditions of pure competition the sellers supply a standardized commodity, so that buyers do not tend to single out individual suppliers. A milling company buying wheat does not care whether it was grown on Smith's farm or Brown's farm, for the products of both farms are equally acceptable and are therefore perfect substitutes for each other.

2. *Monopoly.* When there is only one seller of a product in a market, monopoly exists. Since there is only one, he cannot be substituted for any other seller of *the same product.*
3. *Monopolistic competition.* Where there are several sellers producing articles which are close substitutes for one another, monopolistic competition exists. Monopolistic competition prevails over wide sectors of the economy; it is, in fact, the only form of competition in retailing, and predominates in the service industries. By definition it signifies an admixture of elements of both monopoly and of competition in the sale of commodities and services. For example, a corner drugstore is the only drugstore on the particular corner—it has a monopoly; yet on the next corner there is a second drugstore; the two stores are in competition.
4. *Oligopoly.* The word "oligopoly" itself means "few sellers." Some commodities, such as automobiles and typewriters, are manufactured by only a few corporations, and their competition is characterized by the fact that the actions of each competitor directly and immediately affect the other competitors.

PRICES UNDER CONDITIONS OF PURE COMPETITION

Pure competition exists, then, wherever there is a large number of buyers and sellers of a homogeneous commodity. Pure competition is found in most markets for farm products. For each agricultural commodity there are many thousands of producers, often hundreds of thousands or sometimes even millions. "In 1934 tobacco was grown on 422,000 farms, wheat on 1,364,000, cotton on 1,920,000, and corn on 4,850,000. In 1935 sheep and lambs were raised on 635,000 farms, hogs on 3,971,000, cattle on 5,481,000, and chickens on 5,833,000."[3] No single wheat farmer or cotton grower raises more than an infinitesimally small part of the total wheat or cotton crop. Thus, if a farmer decides to withhold his crop from the market the effect on price would be so small as to be inappreciable; similarly, if a farmer decides to

[3] Clair Wilcox, *Competition and Monopoly in American Industry,* Temporary National Economic Committee, Monograph No. 21, p. 20.

place his entire crop on the market at one time, he is confident that his action will not cause a fall in price.

Under conditions of pure competition the individual seller can sell any amount which he has on hand or can produce without affecting the price: the demand for the product—as it appears to the individual seller—is perfectly elastic. An important consequence of this principle is that there is never any advertising by individual sellers under conditions of pure competition. Advertising is always done to increase sales. If the individual can sell all he wants to, advertising has no value for him. At any market price, the farmer can sell his entire output; at a lower price he can sell no more than his entire output; at price above the market he can sell none at all.

The process of price formation under conditions of pure competition is best understood in terms of the market, short run, and long run, the same time periods used in the analysis of supply. The questions to be answered are these: What determines day-to-day prices? What determines prices when the producers can vary production in the short run in accordance with their expectations of prices? What are the factors influencing the general level of the price in the long run?

Market Price Under Pure Competition

The principal analytical tools for determining market price under conditions of pure competition are the market demand schedule and the market supply schedule. The market demand schedule shows the willingness to buy of the buyers in a market, and the market supply schedule shows the willingness to sell of the sellers; both schedules show different amounts at different possible prices, and both are valid for some one point of time.

Suppose that at 10 A.M. on a particular day on the Chicago wheat market the demand and supply schedules for a certain grade of wheat are those shown in Table 39.

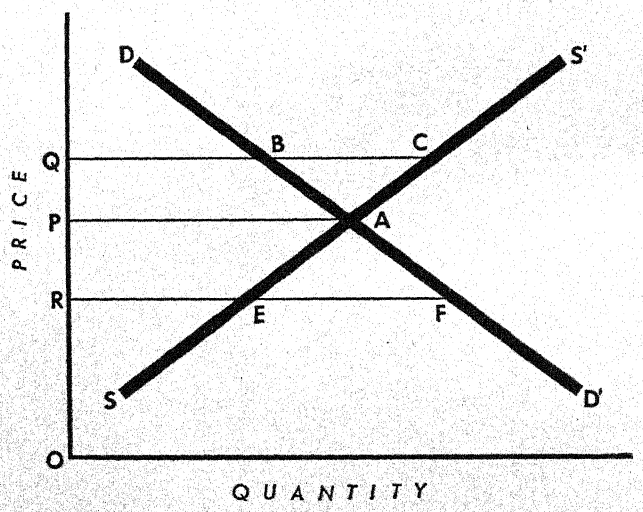
Table 39 **DEMAND AND SUPPLY: MARKET PRICE**

1. Possible prices per bu.	*2. Amount in bu. which would be bought*	*3. Amount in bu. which would be sold*
$1.51	9,500,000	10,500,000
1.50	9,800,000	10,200,000
1.49	10,000,000	10,000,000
1.48	10,500,000	9,800,000
1.47	11,000,000	9,500,000

Columns 1 and 2 in Table 39 are a demand schedule; columns 1 and 3 are a supply schedule. Under the circumstances as posited here, the price established will be $1.49. Could it be higher, perhaps $1.51? No, because at $1.51 the amount the sellers want to sell is larger than the amount the buyers want to buy—the supply *at* $1.51 exceeds the demand *at* $1.51.[4] An excess of supply over demand is also found *at* the price of $1.50. Suppose momentarily that a price of either $1.50 or $1.51 per bushel were established. Then, some of the sellers would be unable to sell all they wanted to sell and thus would offer slightly lower prices, perhaps $1.49½. Their action would then force down the price to $1.49 per bushel. At the price of $1.49, supply equals demand, i.e., the amount which the sellers are willing to sell is equal to the amount the buyers are willing to purchase. Similarly, the price would not be lower than $1.49 per bushel, because if it were demand would exceed supply and competition among the buyers would force the price up.

The price which equates demand and supply is termed the *equilibrium price,* because once established it has no tendency to change so long as the demand and supply schedules themselves remain unaltered. The amount

Chart 13 MARKET PRICE UNDER PURE COMPETITION



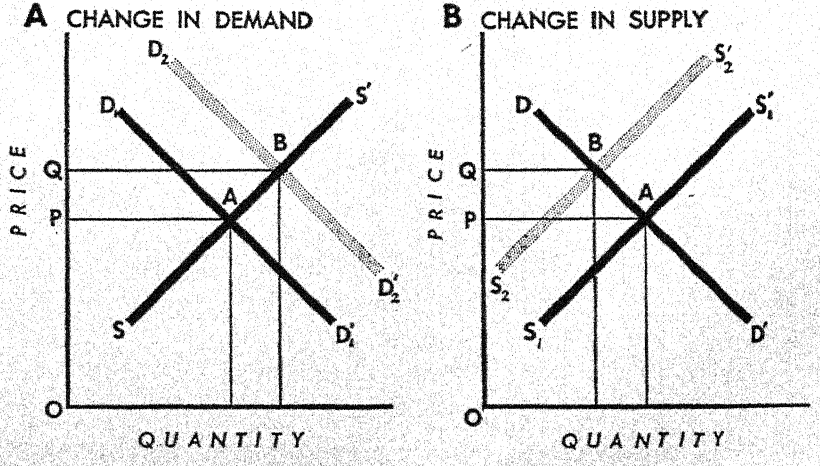
[4] Up to this point the terms *demand* and *supply* have been defined in the *schedule* sense. Frequently in economics the term *demand* or *supply* is used to denote a *particular amount at a particular price,* e.g., at $1.49 the demand for wheat is 10,000,000 bushels. The context should make clear, as in this paragraph, which meaning is intended.

sold at the equilibrium price is the *equilibrium amount*–10,000,000 bushels in this example.

The determination of market price under pure competition is illustrated diagrammatically in Chart 13. Demand is shown by the line *DD'*, and supply by *SS'*. The shapes of the demand and supply curves and their position with respect to one another will vary with the circumstances and expectations of the buyers and sellers. The equilibrium price in Chart 13 is *OP;* the amount *PA* is the amount bought and the amount sold. At a higher price, such as *OQ*, the amount which the sellers want to sell is *QC*, which is larger than the amount–*QB*–which the buyers want to buy. At a lower price, such as *OR*, the amount which would be bought is *RF*, which is greater than the amount which would be sold–*RE*.

Market price, then, is the price which equates market demand and market supply. Now since both market demand and market supply are subject to continual change, it follows that market prices continually change. How are the two kinds of changes related? Suppose that demand increases, while supply does not. An increase in demand, it will be remembered, means an increase in the entire schedule. As a simple illustration, assume the figures in column 2 of Table 39 are each increased by 400,000 bushels. Then the market price would rise to $1.50 per bushel, because at this price both demand and supply are equal at 10,200,000 bushels. In the same manner, the effect of an increase in supply can be shown to lower market

Chart 14 **CHANGES IN MARKET PRICE UNDER PURE COMPETITION**



price. Simple arithmetical examples can be used, in fact, to prove the following propositions:

1. A decrease in demand, if supply is constant, lowers market price.
2. A decrease in supply, if demand is constant, raises market price.
3. If demand and supply both increase, market price will rise or fall, depending on whether the increase in demand is greater or less than the increase in supply.
4. If demand and supply both decrease, market price will rise or fall, depending on whether the decrease in demand is less or greater than the decrease in supply.
5. If demand and supply move in opposite directions, market price will rise or fall, depending on the direction of movement of either demand or supply.

Chart 14 shows the effect on price of changes in demand and supply. In part A of Chart 14 an increase in demand is illustrated; demand increases from $D_1D'_1$ to $D_2D'_2$ and the equilibrium price rises from OP to OQ, the amount bought and sold changing from PA to PB. Or, demand can be imagined as falling from $D_2D'_2$ to $D_1D'_1$, causing a fall in price. Part B of Chart 14 shows either a decrease in supply from $S_1S'_1$ to $S_2S'_2$, or an increase from $S_2S'_2$ to $S_1S'_1$.

Short-Run Prices Under Pure Competition

The short-run price in pure competition is not an actual or market price; instead, it is a *norm* toward which the market price (just described) would tend to move in the absence of other disturbing influences. Short-run price equates short-run supply and short-run demand. It will be remembered that short-run supply is based upon cost schedules under conditions of fixed productive capacity. Short-run demand is a schedule of amounts which would be bought at possible prices upon the assumption that buyers' expectations are definite and unchanging with respect to their incomes, their tastes, and the prices of all commodities other than the one under consideration.

A hypothetical example of price formation in the short run under pure competition is shown in Table 40 which gives the demand and supply for commodity Y.

Under these circumstances the normal price of Y will be 45 cents per unit. Production and demand are balanced, or equal, at the price of 45 cents and 7,000,000 units. Production is at this volume because the producers of Y

Table 40 **DEMAND AND SUPPLY: SHORT RUN**

Possible prices per unit of Y	*Amounts which would be bought*	*Amounts which would be produced*
$0.50	6,000,000	8,000,000
.45	7,000,000	7,000,000
.40	8,000,000	6,000,000
.35	10,000,000	4,000,000
.30	12,000,000	3,000,000

have adjusted their operations to an expected price of 45 cents per unit. If they had expected a price of 50 cents, they would have produced 8,000,000 units, an amount which, however, could not be sold at 50 cents. This amount, it is seen, can be sold only at 40 cents. In other words, if the producers had guessed wrong, their mistake would be corrected by the market. On the other hand, suppose that the price of Y had been underestimated at 40 cents. With this price in prospect only 6,000,000 units would be produced, but this amount would sell for 50 cents. Hence the market corrects this mistake also, the higher price allocating the amount available.

The short-run price, which equilibrates short-run demand and supply from fixed productive capacity, is the only price, of all possible prices, which does not tend to rise or fall so long as the demand and supply schedules remain constant. A price higher than the short-run price will be lowered, because there exists an excess of the amount produced over the demand at the same price; and the excess production can be sold only at a lower price. A price lower than the normal price brings forth a supply smaller than the demand, and it will necessarily be raised.

The short-run price may result in an equilibrium of demand and supply which entails a volume of production representing the maximum of which the industry in question is capable. This will be true in those cases where the demand for the product is very large. Or on the other hand, the equilibrium amount could be a volume of production only one quarter of that possible in the short run. Likewise, the short-run price may be, depending upon the demand schedule, relatively high or low. If it is high enough, the short-run price exceeds the full costs of almost all producers and results in their making large profits. If the price is low it may be below full costs (but above their average variable costs) of all producers and all of them may suffer losses. In other words, the short-run price may or may not cover full costs, and it may cause a large or a small volume of production in relation to the capacity of the industry.

The usefulness of the short-run price analysis may be illustrated by contrasting price movements when (1) changes occur in demand and (2) elasticities of supply of different products vary greatly. The supply of copper, for example, is highly inelastic because, beyond a certain output, low-productivity mines must be counted on for additional production. When demand increases, as it did in 1946, prices rise sharply; production increases also but proportionately much less than price. The low-productivity mines earn their owners only modest returns, but the high price yields the owners of the high-grade mines substantial profits. To take a different example, the cotton textile industry during the 1920's and 1930's had considerable excess productive capacity. Output could be quickly expanded by large amounts with little change in marginal costs. As demand increased only small price increases were required to induce a large expansion of output.

The "Cobweb Theorem"

Although it is generally true that a market price above or below the short-run price is forced by competition back to this norm, there is an important group of cases in which this process does not operate or does so only after a succession of adjustments. These are commonly called cobweb theorem cases (because the diagrams used to illustrate them have the appearance of cobwebs).

The prices of many farm products show definite changes from year to year, or from one- or three-year periods to the next. Sometimes these prices seem to rise and fall rhythmically in cycles—one-year cycles for annual crops, three-year cycles for milk cows. Although such cycles are not rigid and unvarying but are irregular, they may be determined with some accuracy and have indeed been studied for many years. As an example, suppose that growers of a crop based their plantings each year on the price they received in the market in the year before. If the price in Year One was high, Year Two's crop will be a large one. But since a large crops sells at a low price, Year Two's price will be lower than Year One's. Planting for Year Three will therefore be small; and because a small crop sells at a high price, Year Three's price will be higher than Year Two's, and the planting will again be large for Year Four.

In other words, the price in the cobweb theorem fluctuates widely from year to year. It does not return to an equilibrium of demand and supply. However, this state of affairs results only if two things hold true, namely, that production is discontinuous and that producers always base their production decisions solely on last season's price. The first condition is inherent

in the production of most farm products because the seasons make it discontinuous. Once crops are planted, trees set out, herds of young animals started, the pattern of marketable output is set. Crops, of course, can be cultivated or neglected but the amount planted is the determining factor. The second condition of the cobweb theorem is not likely to hold true rigidly. Farmers may base this year's crop on last year's price, but they may not fully believe that last year's price will hold this year and may use it only as a rough guide. A third consideration affecting both of these conditions is that the output of farm products is subject to the vagaries of the weather, labor supply, etc.: farmers' production plans and farm production may be two quite different things.

Prices in the Long Run Under Pure Competition

The long-run price of a commodity is a *norm* toward which the market price would tend to move in the absence of other disturbing influences. As in the case of the short run, factor supplies, tastes, incomes, and technology are assumed to be constant, but unlike the short run the period is sufficiently long for *productive capacity* and demand to adjust completely to each possible price. It is an analysis useful for determining the underlying forces at work upon prices at any particular time, and for ascertaining what ultimate effect these forces would have upon prices if they remained indefinitely in operation without interruption or disturbance.

What is the significance of a price which equates demand and supply in the long run? This question has long been considered important by economists. In the latter part of the eighteenth and the early part of the nineteenth centuries, economic writers devoted much of their analytical powers to the problem of the determinants of the "natural value" of a commodity. Leaving aside the vicissitudes of market fluctuations and other temporary influences, they sought to determine basic economic forces which persist over long periods of time. They were concerned with such real questions as the wages of labor in relation to population, the prices of foodstuffs in relation to the amount of arable land, and the return on capital in relation to both these factors. In modern economics, on the other hand, attention has tended to concentrate on the national income and its development as a determinant of the economic position of a nation. Thus modern economics is less concerned with the "natural values" or "long-run normal prices" of commodities, and is more concerned with the size and composition of the national income. Although the composition of the national income is determined by prices, and although it could be said that the composition

of the national income is affected by adjustments toward long-run prices, attention to long-run normal prices means that fluctuations in national output as well as the effects of technological changes have to be overlooked.

The long run, it will be remembered, is a period in which entrepreneurs can adjust the size of their plants to prices which they expect will prevail for periods long enough to permit such adjustments. It is a period also which is long enough for the number of firms in an industry to grow. Imagine that the demand for a product increases steadily over a period of several years, persisting even through depressions (as did, for example, the demand for mechanical refrigerators in the 1930's), and that the industry producing the product expects a continuing demand. Under these conditions the industry will expand because of the higher profits expected from the rising demand. The industry's long-run supply schedule is shaped by the increase in the number and size of firms and the effect of its expansion on the prices of the factors it uses. Assuming that the technology of an industry remains constant, the schedule for almost any industry will show rising supply prices as the industry grows.[5] (The supply prices are the full costs of production including minimum profits.) Thus, as demand grows, the firms in the industry will be able to produce larger amounts only at rising costs; the long-run normal price, then, will rise as the demand for the commodity expands. This long-run tendency is obscured by changes in technology which lower cost at all levels of output.

PRICES UNDER CONDITIONS OF MONOPOLY

Turn now to the question of how a monopolist determines the price he charges. A monopolist is the sole seller of a commodity or service. (To avoid confusion of terms, note that the word "monopoly" is often loosely employed to designate one business firm's power or dominance even when there are several sellers of a product.) The essence of monopoly is control over prices by the seller. By controlling or setting the price himself, the monopolist determines the volume of his sales and thus also the volume of his own production.

What are examples of monopoly? Before World War II aluminum was manufactured by only one corporation. Electrical energy, telephone service, and gas are almost always sold by only one company in a single territory. The owner of a patent or a copyright possesses a monopoly. Local monopolies are very common. In a small town there may be only one bank, one movie theater, one dentist, one lumber yard, etc., each of which is a

[5] See p. 331.

monopoly. In an isolated farming area a single general store would have a monopoly of retailing.

Although a monopolist is the sole seller of a commodity, the commodity always has substitutes; therefore, a monopolist always faces competition from the sellers of products which are substitutes for his. Oil is a substitute for electric power, steel and copper for aluminum, air mail for the telephone, the mail-order house for the general store. The substitute may not be very close, and the substitute commodity may not be capable of replacing its rival in all uses, but nevertheless its existence exerts a powerful influence in limiting the power of a monopolist.

The existence of substitutes must be kept clearly in mind when examining the problem of monopoly. Candles and kerosene lamps are probably the only important substitutes for electric light in most households. Yet, to the minds of most consumers, they are such poor and unsatisfactory substitutes that the monopoly position of an electric power company is almost unchallenged. On the other hand, the monopoly position of a country general store is not a powerful one. Consequently it is necessary to distinguish sharply between monopoly as such—one seller—and monopoly power—the power to determine prices. The profits of monopoly power depend on demand-cost relations, but they also depend on the power to prevent potential competition.

Monopoly Price

It is well to begin with the pure theory of monopoly price. First of all, for the monopolist as well as for a group of competing sellers, demand is represented by a demand schedule. Suppose that mineral water of certain curative powers is sold by only one firm and that the demand schedule for the product is that shown in Table 41.

Table 41 **DEMAND FOR A PRODUCT OF A MONOPOLIST**

1. Price	*2. Amounts which would be bought*	*3. Total revenue*
$1.00	10,000 gallons	$10,000
.90	15,000	13,500
.80	20,000	16,000
.70	25,000	17,500
.60	30,000	18,000
.50	35,000	17,500
.40	40,000	16,000

Columns 1 and 2 are recognizable as an ordinary demand schedule—here an elastic one. Column 3 shows *total revenue: the total dollar sales at each price* and is obtained by multiplying the figures in columns 1 and 2. It is customary to speak of the "sales schedule" of the monopolist, which is actually the same thing as the demand schedule for his product, although the change in nomenclature shifts the emphasis slightly. The emphasis is on *sales possibilities* of the monopoly, i.e., on the demand schedule which, in a way, he can manipulate.

At what price will the monopolist sell? To answer this question one must know his costs as well as his demand. Suppose simply that his costs are zero; e.g., that the mineral water is sold directly from a spring. In such a case it would pay him best to set a price of 60 cents for a gallon of mineral water. This price maximizes his total revenue. A higher or a lower price would mean a smaller total revenue.

Note that the monopolist sets a price which maximizes his revenue unless for some reason his aim is not maximum profit. Under the temporary assumption that his costs are zero, his net revenue is also maximized. A monopolist does not set the highest price he can charge, nor the price which would yield the greatest profit per unit: *monopoly price is one which maximizes total profit.*

When costs are introduced, the principle is the same. Profit is the difference between total costs and total revenue, and the monopoly price is one which maximizes this difference.

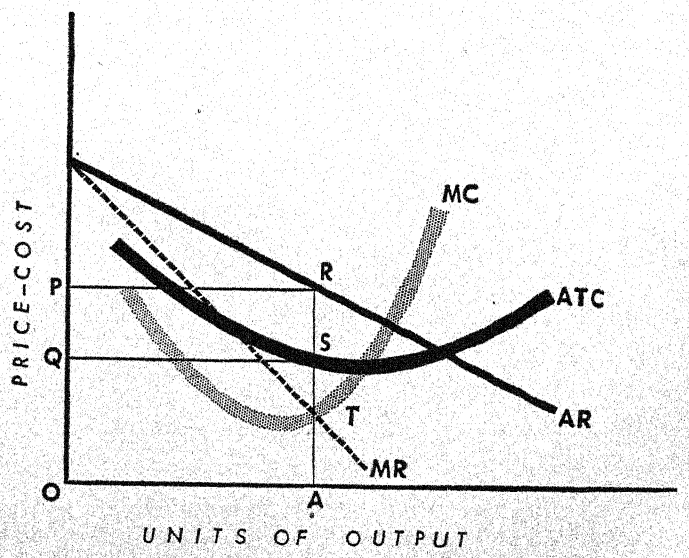
A graphic illustration of the principle of monopoly price is given in Chart 15. In this chart the line *AR* is the average revenue curve: it represents the sales possibilities of the monopolist, and is also called his sales or demand curve. The line *MR* is the marginal revenue curve. Marginal revenue is the addition to total revenue derived from the sale of one more unit of the commodity. Marginal revenue falls faster than average revenue or price, because as a monopolist increases his sales and therefore must accept a lower price, marginal revenue takes into account not only the fall in price but also the loss in revenue on all previous units as additional units are sold at a lower price. The curves *ATC* and *MC* in Chart 15 are the average total costs and the marginal costs, respectively, of the monopolist.

The monopolist will maximize his profits when he sets the price at *OP*. At this price he produces and sells the amount *OA*. The marginal cost and the marginal revenue of the amount *OA* are equal at *AT*. The monopolist's profits are the difference between his total revenue, which is his sales *OA* multiplied by price *OP* (and therefore the rectangle *OARP*), and his total costs, which is his sales *OA* multiplied by average total cost *AS* (and there-

fore the rectangle *OASQ*). His profits are hence the rectangle *QSRP*. These profits are a maximum because if the monopolist produced and sold more than *OA*, the graph shows that marginal cost—the addition to total cost—would exceed marginal revenue—the addition to total revenue, and thus profits would be smaller. Similarly, profits would be smaller if output were lower than *OA*, because marginal revenue then would exceed marginal cost, and profits could be increased by expanding output to *OA*.

Monopolies on occasion have resorted to the physical destruction of commodities: they have dumped goods into the sea, burned them, or otherwise destroyed them. They have done this presumably to make their profits larger. But how can it be profitable first to produce commodities in order later to destroy them? The deliberate destruction of commodities is usually an attempt to limit supply to attain a favorable price. For example: fish are caught by several hundred fishermen who pool their catches in a co-operative marketing organization which happens to be a monopoly. The monopoly sets a price which it believes will maximize the total revenue; but the amount which the monopoly thinks can be sold at this price is less than the combined catch of the member fishermen. The monopoly organization then orders the destruction of the "surplus" fish. Of course, difficulties of storage affect such a decision; furthermore in such an organization control over the members is not too strict; and to prevent any member from

Chart 15 **MONOPOLY PRICE**



breaking the price by surreptitiously selling more, the surplus product must be destroyed.

The physical destruction of commodities by a monopoly illustrates two important points. First, where a relatively limited amount sold on the market brings a larger total revenue, demand is inelastic. If demand were elastic, it would pay to sell all that could possibly be sold, however low the price, because total returns would be larger. Destruction, then, can be a deliberate policy only if demand for the monopolized product is inelastic or is believed to be so. Second, destruction is of necessity a market or perhaps a short-run policy; it is never a long-run policy. Goods of any kind involve production costs, however large or small. No one in his senses would plan to produce something and then to destroy it. Thus destruction by a monopoly occurs only when a mistake is made, when either anticipated demand is overestimated when production is carried out, or when a monopoly organization fails to prohibit its members from producing "too much."

Price Discrimination by a Monopolist

Since a monopolist is free—within the limits of the demand for his product and the limits of his costs—to choose a price for his product, it follows that he can select two or three or more different prices if he finds such action expedient. Thus, for example, an electric power company charges one rate for electricity for domestic consumption, another rate for electricity for advertising signs, still another rate for electricity for industrial power, etc. *The practice of charging different prices for the same thing or substantially the same thing is commonly called price discrimination.* Price differentiation is probably a preferable expression, since it avoids any derogatory connotation given by the word "discrimination" in this context. The practice is common; professional men employ it as a matter of course. A surgeon charges one patient one thousand dollars and another only fifty dollars for identical operations; lawyers' fees sometimes vary as greatly. The Federal government charges different rates for different classes of mail.

Price discrimination is practiced by a monopolist because it increases his profits. Ideally a monopolist's maximum profit would be obtained if he could charge each one of his customers the maximum amount each individually would pay for the commodity or service, in other words, a different price for each customer (unless some of them had identical individual demand schedules). Actually only a rare monopolist, one who has a very

limited number of buyers whose tastes and incomes are known, can practice such refined discrimination. What the monopolist can do is to divide his market into a few classes or groups and to charge each group a different price. For example, a monopolist may sell at one price to the domestic market and at another price in foreign markets. Railways in Europe sold their passenger services at three prices: first, second, and third class. In the United States, they have two prices, Pullman and coach. Theater owners divide their seats into groups—orchestra, balcony, etc., at different prices.

In practicing price discrimination the monopolist must be able to maintain the distinctions he makes in the market, because otherwise no one would pay any price above the lowest, or the customers who could buy at lower prices would resell the commodity to those buyers who are expected to pay the higher prices. Distinctions between the buyers from a monopoly can be made and enforced in many ways. Where a monopoly sells at a lower price in the foreign than in the domestic market, a tariff wall keeps domestic consumers from buying the product abroad and importing it. Ticket arrangements and supervision by conductors or ushers keep the people who buy coach or balcony tickets from sitting in the Pullman or loge seats. Technical barriers are sufficient to prevent the householder from buying electricity at the industrial rate. A glass monopoly charges more per square foot for large pieces than for small ones: the buyers cannot make big ones out of little ones.

Elasticity of demand in each of the separate markets in which he sells his product determines the difference between prices charged by a monopolist. The price is higher in markets where demand is less elastic and lower where demand is more elastic. Suppose the owner of a mineral spring puts out his product under two brands. The mineral water is the same, but the two names, two labels, two shapes of bottle give the unwary the impression that they are different. One of the brands is marketed through hotels and restaurants and the other through a grocery chain; one brand is thus sold to well-to-do buyers, the other to middle-or-lower-income buyers. The demand of the well-to-do buyers is less elastic, because the ratio of any possible price to their incomes is low; therefore the price can be raised without diminishing the sales greatly. On the other hand, the customers of the grocery store get the second brand at a much lower price, because their demand is elastic, and they will buy more at the lower price. By this discrimination total profits are greater than if the one brand were sold at a uniform price.

Qualifications to the Principle of Monopoly Price

So far it has been assumed that a monopolist charges the price which maximizes his profits. This assumption is not valid for all monopolists because, among other reasons, a monopolist (1) may fail to know adequately the market (demand schedule) for his product and thus the optimum price, (2) may be satisfied with present profits, (3) may fail to experiment with different prices to determine optimum prices because of inertia or some other inhibition. The price fixed by a monopolist, then, can be different from the profit-maximizing price.

Another important reason for establishing a monopoly price different from the optimum is fear of potential competition. In such cases, the price is below the optimum position and profits are less than the maximum. This circumstance applies usually to those monopolists whose position is tenuous, particularly to local monopolies. A lower than optimum price may yield profits considered adequate by the monopolist, yet not so high as to induce others to set up competing businesses.

The prices of the most important groups of monopolies, the public-utility corporations, are of course regulated. The effectiveness of regulation varies from case to case, but in general the prices charged by public utilities are fixed by regulating bodies at a "competitive" level, which is taken to mean that prices are equal to "costs plus a fair profit." Rate-fixing legislation was prompted by popular outcries against the high prices formerly charged by monopolies. At the present time other monopolies which have a public-utility status, either not subject to rate regulation or subject only to inadequate rate regulation, may charge a lower-than-optimum price because they fear that a high (profit-maximizing) price would lead to similar imposition of effective price regulation.

The Significance of Monopoly

Under monopoly, prices tend to be higher and production tends to be lower than they would be if the costs were the same and the commodities were sold under pure competition or even under monopolistic competition.

The relatively higher price charged by a monopolist is made possible because he can choose from among all possible prices that one which maximizes his profits. A seller under competitive conditions also tries to maximize his profits, but his power to select an advantageous price is circumscribed by rivals. Under pure competition he does not select a price

at all but must adjust himself to prices over which he has no control; under monopolistic competition, the seller has only a very limited power to set a price. By charging higher than the price which would exist under competitive conditions, the monopolist therefore "exploits" his customers. Exploitation varies, of course, and is less where a monopolist, because of ignorance, inertia, or fear of potential competition or regulation, charges less than he could.

Since a monopolist's price is higher, it follows that the amounts produced and sold are less under monopoly than they would be if the product were produced at the same cost and sold under pure competition. Restriction of output thus characterizes monopoly. Restriction of output has several important consequences: consumers get less of the monopolized product than they are willing to buy at a price which covers full costs together with a minimum profit. It follows also that the resources employed by the monopolist are less than the resources the consumers "want" to be employed. The allocation of resources in the economy as a whole is distorted through the actions of monopolists in producing less than they could. Before World War II the aluminum monopoly, for example, caused the price of aluminum to be high relative to other competitively sold materials. Hence less aluminum and more of other materials were used than if all materials had been competitively sold. The amount of capital and labor employed in aluminum production was too small, and that employed in the production of substitutes for aluminum was too large.

What are the favorable aspects of monopoly? They are indeed few. A common argument supporting the existence of monopoly runs about as follows: monopoly arises because modern technology necessitates large-scale production, which in turn entails concentration of production in one firm which can produce at very low costs. In other words, monopoly is justified on the grounds that it encourages the full use of technological advancement and that it makes possible very low costs of production. Socialists as well as capitalistic spokesmen of monopolies themselves are very fond of this line of argument. It has, however, only scant merit. There is no general tendency to large-scale production in the economy. Differences in cost between large and small business units exist, but empirical studies have shown that they are not great. Moreover, the large corporation controlling many plants must not be confused with the plant itself. The former is often large, but there is nothing technologically necessary about such size. A monopoly which points to the decline in its prices over a twenty-year period is not proving the efficiency of monopoly; it may only

be showing that technology has improved or that the monopoly has taken advantage of the economies of size. The burden of proof lies on the monopoly to show why its prices should not be still lower.

PRICES UNDER CONDITIONS OF MONOPOLISTIC COMPETITION

Monopolistic competition, as already noted, is the term applied to the competition among sellers of products which are unique yet close substitutes for one another. It exists in many forms: price competition, quality competition, advertising competition, etc. To indicate the multitude of different kinds of monopolistic competition, a partial list of examples of nonprice competition follows.[6] Each of the items in the list represents a concession which a seller can make to a buyer, and which therefore gives a competitive advantage to the seller unless his rivals make the same concession.

1. Concessions with respect to time of buyer's payment:
   - Discount
   - Credit terms
   - Installment sales
   - Deferred payment
2. Concessions related to risks of buyer:
   - Guarantees
   - Price guarantees
   - Accepting return of obsolete, discontinued, or "unsaleable" merchandise
   - Exchanging merchandise
   - Sales subject to trial
   - "Money-back" agreements
3. Concessions related to supplying additional goods:
   - Premiums
   - Combination sales and offers
   - Samples
   - Special containers
4. Concessions rendered buyer through use of seller's employees or property:
   - Demonstrating
   - Furnishing plans and specifications
   - Installation and erection
   - Repair and maintenance
   - Reconditioning
   - Delivery service by seller's or agents' trucks

[6] Adapted from *Price Behavior and Business Policy,* Monograph No. 1, Temporary National Economic Committee, 1940, pp. 94-98.

5. Concessions rendered buyer through financial assistance or favors:
   - Entertainment
   - Patronizing publications in which buyer is interested
   - Paying buyer's personal expenses
   - Assuming reversed telephone or telegraph charges
   - Assisting buyer to obtain used products for trade-ins
6. Concessions through payment or diversion of commissions or fees to buyer:
   - Payment of commissions or fees by members to purchasing agents compensated by buyers
   - Splitting of commissions or fees by agents of seller with agents of buyer without buyer's knowledge
7. Concessions through allowances or payments for value rendered by buyer:
   - Trade-in allowances
   - Advertising allowances
   - Container allowances
   - Purchasing from buyer
   - Renting from buyer
8. Concessions granted during performance contrary to provisions of agreement:
   - Settlement of old accounts at less than full value
   - Permitting buyer's cancellation of order
   - Substitution of higher quality of goods

The items in this partial list are not only methods of competition, but also methods of modifying price quotations. Thus, for example, when a used article is traded in on the purchase of a new one, the trade-in allowance modifies the price. Two dealers may have the same price, but if one offers a higher trade-in allowance, his price is in effect a lower one.

Sometimes sellers wish to conceal the concessions which they make to buyers for reasons of competitive advantage. Devices of concealment are likewise many in number, as may be noted from the following partial list:

1. Acceptance of forms of payment in which concessions may be concealed:
   - Accepting securities
   - Accepting buyer's capital stock
2. Types of agreements, offers, invoicing, etc., by means of which concessions may be concealed:
   - Oral agreements
   - False billing or receipts
   - Misdated invoices or contracts
   - Invoices omitting terms of sale
   - Unitemized billing
   - Orders omitting specifications

These competitive practices display the complexity of monopolistic competition. To reduce the complexity, certain general concepts are necessary. One of these is *product differentiation.*

Product differentiation signifies the fact that competition in wide areas of the economy is carried on by sellers of similar but differentiated products. The automobile industry, for example, produces automobiles which may be thought of as a distinct commodity, but actually automobiles consist of Chevrolets, Fords, Plymouths, and other makes, all of which are distinguishable from one another. Each is unique. It may be seen in this sense that Ford has a monopoly of selling Ford cars, General Motors of Chevrolet, and Chrysler of Plymouth. That "monopolies" of this kind are not unreal is borne out by the well-known fact that some buyers always purchase the same make of car and swear that they would never shift to another. Similarly, each retail store has customers who never buy elsewhere. In general, each seller under monopolistic competition has a portion, large or small, of his total customers who will always buy from him. Thus the individual seller under monopolistic competition has his own market, as well as a "shared market" with his competitors.

In order to build up a market of his own the individual seller goes to great lengths to differentiate his own product from those of competitors. A distinctive name, carefully chosen packaging or labeling, unique design, and above all advertising are designed to impress the consumer with the special qualities of the product. The product itself, its brand name, its label, and its advertising are protected by the patent and copyright laws. Some manufacturers have been so successful in their efforts to impress brand names on the buying public that these brand names have become common generic terms for commodities; e.g., kodak, linoleum, kleenex, simonize, scotch tape, frigidaire, and dictaphone.

While product differentiation causes the buyer to distinguish one brand from another, the seller's attempt to make his product different cannot be extended too far. Product differentiation actually involves a considerable amount of imitation. Where is the best place to open a new five-and-ten-cent store? Nearly always, it seems, the best place is near an established one. New five-and-ten-cent stores, furthermore, tend to display the red and gold lettering of the older established stores of the same kind. A new manufacturer of automobiles does best if his car has the general qualities of the existing makes. If the new car is radically different—if product differentiation is extreme—consumers may be wary. These are illustrations of the principle that a new business or a new product often will have the most

favorable prospects for success if it resembles established ones sufficiently to attract part of the established market.

The imitation which accompanies product differentiation accounts for the shrillness of so much advertising: several brands of a commodity may be all very much alike and are made alike because each manufacturer tries to maximize the potential number of consumers, but to draw the consumer to himself each manufacturer emphasizes a trivial point in his advertising. In fact, the owner of a well-known brand may be viewed, for some purposes, as being in a clear-cut monopoly position. Such products as Coca-Cola, Cream of Wheat, and Simonize do face competition from products bearing other names, but so well established are these brands that their owners face much less competition than the owners of the lesser known brands.

Problems of the Seller

The individual business unit under conditions of pure competition has a relatively simple task: to adjust costs and output to a price which he is unable to control. Monopolistic competition is more complicated because the seller has to adjust price as well as costs and output, and corollary factors are his advertising budget, his sales methods, and other means of maintaining competition. The problem of the seller can be grouped under three headings: (1) price policy, (2) quality policy, (3) sales policy.

Prices Under Conditions of Monopolistic Competition

How are the prices of commodities determined under conditions of monopolistic competition? When discussing a *commodity* in this context, it must not be forgotten that the commodity itself has no one price. Each seller's product is differentiated from the others and may have its own price. Thus if there were two hundred brands of a canned vegetable, there could be two hundred separate prices, although in practice many different brands will sell at the same price. The problem of visualizing prices under monopolistic competition is twofold: (1) the question of what decides the price a particular seller will select; and (2) the question of what determines the cluster of prices of different brands of the same commodity.

The price charged by an individual seller depends on the shape and position of the demand curve for his own product in relation to his costs, which means that his problem in fixing a price is basically the same as the monopolist's. He, too, attempts to maximize his profits. His position differs

from the monopolist's, however, in that the demand for his product is highly elastic because there exist close substitutes for it. A high price would mean a very small volume of sales because buyers would turn to rival brands. Furthermore, if profits are large the seller under monopolistic competition can expect fresh competition from businesses coming into his field. The new competitors capture part of the market, thus diminishing the demand and hence the profits of the enterprise already in the field.

Variations in the prices charged for different brands of the same commodity exist largely because of variations in cost among the producers and in the demands for the several brands. Well-known and firmly established brands will have large demands which are less elastic; less-known brands will have smaller demands which are more elastic. This means that the enterprises selling the well-known brands can often charge higher prices than the others. Some of the well-established brands may be hardly affected by competition from new brands, and hence profits of the well-established firms may be reduced only to a slight extent by the entry of new firms into the field. It follows that the prices of the several brands of a commodity may continue at different levels over long periods of time.

The level of the whole cluster of prices of the brands of a commodity depends on the relations of the prices to volume of sales and to costs. If the relations are such as to produce large profits for most of the firms in the industry, competition from new firms attracted by the high profits will result and act in such a way as to lower prices. In general, then, the cluster of prices is held down except temporarily to the levels of costs by the force of competition. This is true, however, only if there is perfectly free entry to the industry. If profits are large in an industry, perhaps owing to an increase in demand, new firms enter the industry unevenly: they set up brands of their own, but the new brands compete unevenly, i.e., with different degrees of success, with the established ones.

Product Variation

Monopolistic competition is not always on a price basis. Competition is more often based on differences, or seeming differences, of quality rather than of price between products. In fact, it is a major characteristic of monopolistic competition that some commodities maintain the same price over several decades. Some magazines, for example, may go through their entire life cycle—birth, growth, decay, and suspension—at the same price; others may change their prices only under severe circumstances, as when during World War II magazine prices went up from 5 to 10 or 10 to 15

cents. Competition is keen between news and picture magazines, digests, etc., but it is centered on differences in content and format rather than price, with the sole and not very important exception that sometimes subscriptions are offered to "selected" groups of persons at rates which amount to a reduction in price. The problem of the individual seller is to adopt that degree or form of quality in his product which maximizes his profits. Stated very simply, it is a problem of effecting a happy compromise between the smaller costs of a lower quality and the larger sales of a higher quality. Quality is here to be understood as those features of a commodity which seem valuable to the consumer. If the consumer is impressed by a glittering exterior, an extra gadget, or some other superficial aspect of a commodity, the manufacturer may find that he can increase his sales greatly by spending a little more on his product, so that his profits may increase almost as much as his sales. On the other hand, it may be profitable to worsen the materials or workmanship in a product for this may go unnoticed by all save a few of the buyers, with the result that costs fall and profits rise.

For some commodities there are price lines, i.e., a list of prices at which the various grades of the commodity are sold. In the case of women's clothing, there is a series of definite prices, established by the accumulation of habit and tradition. "Prices have come to be accepted as virtually immutable, except that new lines may be added when extreme changes in cost occur. They are rarely, if ever, considered subject to modification by bargaining. In planning production, the manufacturer carefully estimates the material and labor costs which will be required to produce a desired model and then makes such modification in the design as will be necessary to permit him to sell the product at one of the accepted lines." [7]

Product variation may take the form of changes in quantity as well as in quality. Table 42 shows how a producer actually changed the size of candy bars selling for 5 and 10 cents.[8]

Table 42 **CHANGES IN WEIGHTS OF CANDY BARS**

Date of change	*Net weight per bar (plain)*	
	5 cents retail	*10 cents retail*
Jan. 1, 1926	1⅜ ounces	2¾ ounces
Oct. 10, 1931	2	4
Sept. 2, 1932	1⅞	4
Nov. 4, 1932	1⅞	3¾

[7] TNEC Monograph No. 1, p. 70.

[8] TNEC Monograph, *op. cit.*, p. 73.

What are the consequences of product variation, of quality competition? Does quality tend to be high or low under monopolistic competition? One thing is certain: monopolistic competition in the American economy has provided the consumers with a luxuriance of choices. In many cases monopolistic competition has resulted in a steady rise of quality, although it would be difficult to say how much influence to ascribe to monopolistic competition on the one hand and to improved technology and the general upward trend of the national income on the other. No seller can afford to let the quality of his product drop very far below the levels set by his competitors. Where monopolistic elements are strong in a nominally competitive situation, quality deterioration often exists. As in price, so in quality a monopolist arranges matters to suit himself, subject always to the nature of the demand for his product.

Monopolistic Competition and Advertising

Advertising is a means, like lowering prices or improving the quality, by which a seller tries to improve sales under monopolistic competition. Advertising is the leading example of the methods by which the seller seeks to increase the demand for his product. Other methods include the establishment and maintenance of adequate retail outlets, by expenditures on the commissions and expenses of salesmen, by providing margins for retailers, by providing retailers with display material, and the like. Advertising is the principal selling cost, as was previously noted, and for the sake of illustration advertising may here be considered as synonymous with all selling costs.

Advertising is intended to increase demand. Herein is one of the sharpest distinctions between pure and monopolistic competition. Under pure competition there can never be any reason for the *individual* seller to increase the demand for his own product. At the prevailing price, the demand for his product is infinite (he can sell any amount he produces without affecting price); and since his product is identical with that of other producers, there is no reason for attempting to increase demand. Under monopolistic competition, however, the individual businessman has to sell his brand in competition with others. If he has a product which is priced competitively and the quality of which is likewise set up on a competitive basis, it may be that the demand for his brand may be very small; few consumers know of it and there are many similar brands. He decides to incur the expense of advertising to increase demand. The effect of advertising to the seller is thus similar to improved quality: higher costs are incurred in order

to bring in relatively more revenue. Advertising also resembles a reduction in price: increased sales, but a gain in revenue of smaller proportions. Hence the first consideration is always the relation between advertising expenditures and the additional demand they create. In general, it is true that advertising is subject to diminishing returns; i.e., there is some point beyond which it does not pay to increase advertising, because the additional revenue from sales is less than the cost of the advertising itself. Consequently, each seller in planning an advertising program must, on the basis of experience or otherwise, decide just how far to carry his expenditures in advertising.

Although advertising is always intended to increase the demand for the thing advertised, it sometimes increases the demand for rival products as well. A buyer who is attracted by an advertisement may in a sense discover a new want; and after consideration he may buy the commodity, which is new to him, but he possibly buys a brand competing with the advertiser's. This is the reason why advertisements admonish the buyer to accept only the "genuine, to take no substitutes," to be sure to look for the "red and beige label," and to remember "if it hasn't the blank mark, it isn't Blank's." The effect of advertising, then, is not predictable, because the benefits of the advertiser can be accompanied by benefits to his rivals, whose position he would prefer to deteriorate rather than to improve.

Considered in isolation, there is some optimum of advertising expense for an individual seller, a specified amount of advertising which will make his profits greatest. But, considered in a competitive context, the matter is different. Suppose that one firm in an industry increases its advertising budget considerably and enjoys an increase of sales. The new sales can come from wholly new buyers or by a diversion of buyers from the other firms in the industry. These latter firms, seeing the profits or at least indications of profits in the former and perhaps suffering losses of their sales, also begin to increase their advertising heavily. The effect is, perhaps, that total sales of the industry are somewhat larger, but each firm shares in this so that each has only a slight increase in sales. A further possibility is that now all firms are worse off than before the general increase in advertising, and that none dares individually to diminish his outlay in advertisements. Another possibility, and a more likely one, is that the firm which initiated the increase retains its advertising advantage over the others and its larger profits.

As a competitive weapon, advertising itself may be used in many ways, and it is but one of many other devices for increasing demand. It is impossible here to describe all the complexities of selling effort, but a few

important generalizations can be made. Advertising is the principal method of competition in the sale of some products. Price and quality are often about the same for different makes and brands, but some sellers are much more successful because of advertising.

Both prices and quality are affected by advertising. In general, prices of advertised goods are higher because of the advertising itself. Advertising is a cost and where prices cover costs prices must be higher in proportion to advertising expenses. This statement is certainly contrary to the popular opinion that advertising makes possible "mass markets," and therefore mass production and the low prices that are believed to result from mass production. This opinion is indeed valid for many individual cases. The demand which can be created by advertising does mean lower unit costs of production (including the advertising)—*but only in those cases where there are actual and substantial economies of large-scale production.* But economies of size are not universal, and where they do not exist advertising causes prices to be higher. However, a well-advertised branded product may be handled by distributors on a lower margin of profit, i.e., a lower gross margin. Thus the retail price is not higher since the retailer himself bears part of the manufacturer's advertising costs and does so because his customers ask for the well-known products.

Since advertising is a substitute method of competition for quality improvement, it follows that advertising must sometimes replace a quality improvement that would otherwise be made. On the other hand, advertising serves to maintain quality standards, because faith in the advertising is considered to depend on assurance of given quality.

CHAPTER 16

Prices and Markets: II

In examining the formation of prices thus far the conditions of pure competition, monopoly, and monopolistic competition have been described. Another condition of great importance is *oligopoly,* in which *the number of*

sellers is few enough so that each may directly affect the price of a commodity by means of his supply of it.

OLIGOPOLY

Oligopoly exists in many branches of American industry: aluminum, steel, automobiles, farm implements, plate glass, cigarettes, and typewriters are produced in industries which consist of two ("duopoly"), or three, or a few more companies. In some industries two or three or four companies produce and sell nearly the entire output of a commodity, even though there may be a dozen other producers, the main fact being that very few large companies dominate the entire industry. There are, then, two important aspects of oligopoly: (1) fewness of sellers, and (2) differences in size among the few sellers.

As among monopolies, local oligopolies are quite common. In small cities and in certain districts of large cities there are always two or three drugstores, banks, hardware stores, etc. Competition in such markets is characterized by the fact that each of the sellers must take direct account of the actions of the others, and also calculate the effect of his actions on them. Oligopoly is one form of *competition.* This point must be emphasized strongly to avoid confusion. Suppose that in an industry composed of five independent corporations there is a tacit or an explicit agreement on prices, on sales and sales policies, and on standards of production: this is monopoly. Suppose these five corporations are actual rivals, each seeking to increase its profits at the expense of the others: this is oligopoly. The existence of only a few sellers in an industry or in a locality is not, therefore, proof of oligopoly. Under present law it is illegal for sellers of a commodity or service to unite their policies in such a way as to create a monopolistic condition. The Sherman Act of 1890 and the Clayton Act of 1914 prohibit agreements in restraint of trade, i.e., monopolistic agreements. Monopolistic practices, then, exist only if enforcement of these laws is weak or if these laws as interpreted by the courts do not cover some forms of agreement. It is obvious that the fact that oligopolists follow the same course of action is not necessarily an indication of explicit agreement among them. Several firms in an industry may sell at the same price because each has arrived independently at the conclusion that this price is the most profitable despite individual differences in costs.

Prices Under Oligopoly

The problems of price formation under conditions of oligopoly are well illustrated by *duopoly*, which is a form of oligopoly limited to two sellers.

In duopoly the salient fact is the high degree of interaction in pricing between the two firms. Suppose two railroads, A and B, serve a common territory with parallel lines and that they are not subject to government regulation of their fares and rates. If A lowers its rates (B's remaining the same) it will divert to itself most if not all of the traffic served by B; and if railroad B should in turn cut its rates below A's, a similar diversion of traffic would happen. Price-cutting would be advantageous to a point since the marginal costs of railroads are low, i.e., the extra cost of serving extra business is relatively small. Price-cutting by railroads could be extended very far before the additional revenue would be less than the additional costs. One possibility under duopoly, then, is drastic price competition or cutthroat competition, as it has often been called. This is about what happened among railroads in the period before the establishment of regulation by the Interstate Commerce Commission in 1887. Rates were sometimes cut to extremely low figures when two railroads were engaged in direct price competition. On the other hand, railroads in the same region often agreed on rates, setting them high, thus establishing temporary monopolies.

The principle illustrated may now be stated: the demand schedule for the product of a duopolist depends on his price and on the price charged by his rival. The total demand schedule for the products of both sellers depends, of course, on the buyers' tastes and incomes, and the prices of other commodities and services. But the demand schedule for the product of one of the rivals is a different thing. Knowing that a reduction in price is certain to be followed by retaliation, under these conditions each seller will seek to set a price which will maximize his profits but which will give his rival no recourse for retaliation. If both sellers (or several sellers) had the same costs and if the product were undifferentiated, each would set the same price. The price would be identical with the price fixed by a monopolist under the same circumstances. This conclusion depends on the assumption that each seller considers the effect of his own price policy upon his rival and the rival's possible countermeasures, that he knows the total demand schedule for the product, and that he attempts to maximize his profits.

Thus competition which takes the form of rivalry between a few sellers, oligopoly, can produce the same consequences as monopoly. The broader

implications of monopoly have been noted: in general higher prices, lower outputs, and less economic use of resources than under pure competition. These consequences also follow from some oligopolies.

Oligopoly, like any other general form of competition, manifests itself in a variety of ways. Thus far in this analysis it has been assumed that equal costs, perfect knowledge of the market, and profit maximization characterize each competitor. Moreover, product differentiation has not been introduced as a factor in oligopoly. By altering these assumptions and by combining them in different ways, several typical oligopolistic situations become evident.

Suppose that the costs of the several sellers of a product are unequal. If the oligopolists are about the same size, the seller with the lowest costs will set a price which the others will be forced to follow if they desire to maintain their sales. The lowest cost seller, under these circumstances, can act as if he were a monopolist, estimating demand, price, and his own cost so as to maximize his profits. This process is termed *price leadership*. Price leadership arises also where differences in size are important. One firm may have 50 per cent of the sales of a product, its several competitors the remainder. This proportion can come about when the large firm has been in the field for some time, the others being relative newcomers. The large firm sets a price which the others follow even if the large firm has higher costs than the smaller ones. Its position of price leadership is possible because its powers of retaliation are very great: it can afford to undersell a small competitor for a period long enough to ruin him.

Product differentiation under oligopoly makes possible the existence of various prices at the same time for different brands of a commodity. If there were no product differentiation under oligopoly—e.g., two parallel railroads offering identical services—the sellers would have to set the same prices if they are to survive. That is to say, the price of homogeneous products sold under oligopolistic conditions are directly tied together, because the only form of competition is price competition. With product differentiation, prices can be different. One seller can emphasize the high quality of his product and thus charge a higher price than the others, while the other sellers could produce medium and lower qualities at lower prices. Although prices may not be identical, they must be "in line" (i.e., price policies must be compatible) if the industry is not to be subjected to price wars. Since price wars can be destructive as well as nerve-racking, it is easy to understand that the inclination to engage in price-cutting—in price competition—is not a strong one.

Price competition is in a sense the fundamental form of competition.

Under pure competition, it is prices which adjust demand to supply and supply to demand, and which decide who shall produce and who shall buy and who shall not. Yet price competition is uncommon in oligopoly as also under monopolistic competition.

The decline in price competition in the last few decades is to be attributed partly to the growth of oligopoly, which is a consequence of the concentration of production as well as of corporate control. But in addition the attitudes of American businessmen have changed. Price wars were not rare in the latter part of the nineteenth century. They were frequent among railroads before the imposition of federal control; and they characterized the growth of the oil and other industries. In the period since World War I, however, price competition in oligopolistic industries has been considerably attentuated. Price reductions as a method of gaining a direct competitive advantage came to be looked upon almost as unfair competition. It was held that competition ought to be in quality, performance, and service. Trade associations and other business organizations have campaigned against the idea of direct price competition and advanced price stabilization in its place. During the 1930's, in fact, price-cutting became a term almost of opprobrium; a "price chiseler" was condemned.

The "rigidity"—the infrequence and small extent of change—of prices under oligopoly is thus attributable to the prevalence of nonprice competition. Specifically, price rigidity can also be explained in terms of the demand for his product as it appears to the individual oligopolistic seller. Suppose he considers raising his price. This will be unwise if his rivals do not follow, because many of his customers will desert him if he raises his price: he visualizes the demand for his product as being very elastic above his price. Again, suppose he considers lowering his price. In this instance he may believe that his rivals will follow him and cause his gain to be small: he sees the demand for his product at lower prices as being inelastic. If such are a seller's opinions, small wonder that he finds no reason to change a price that is clearly not unsatisfactory. And if his rivals are of the same mind, the price structure in the entire industry will be rigid.

Price changes, of course, do take place under oligopoly, but they are likely to be the direct and appreciable consequences of specific causes, such as wage increases, tax changes, cost reductions, new processes, technological improvement, etc. Price reductions which represent a reduction in cost are not in the same disfavor as price cuts which are purely competitive measures. If a corporation lowers its prices after effecting economies in manufacturing, the other corporations in the industry may be impelled to follow whether or not they have been able to reduce their

own costs. The fact that they will follow could be a deterrent to the progressive firm which nevertheless may be sufficiently confident of its prospects to proceed with a price reduction anyway.

PATTERNS OF OLIGOPOLY PRICES

Since under oligopoly the possibilities of price competition are so great and since price competition where there are only a few sellers can be dangerous, many industries have established means to avoid price competition and in some degree to control prices. Among others, these are price leadership, zone price and basing-point systems, and freight equalization systems.

Price leadership, which has already been described, is found wherever all or most of the firms in an industry adopt the price set by one firm, or adjust their prices in a definite relation to the price of the leading firm. Price leadership has existed in the steel, petroleum, fertilizer, industrial alcohol, and other industries, the price leader usually being the largest firm in an industry. Firms follow the price increases or decreases of the leader because, for one thing, they may be afraid to compete on a price basis with the largest and most powerful company in the industry. For another, they may genuinely respect the policies of the big firm. After all, it is no simple matter to determine a price in view of the complexities of cost and market analysis, and to copy the price schedule of a successful firm is safe and perhaps profitable. Then, too, the industry may believe in the desirability of stable prices, and because outright price agreements are beyond the pale of the law, price leadership becomes a kind of agreement by example. Finally, price leadership is to be expected where there are great differences in size among the firms under oligopolistic conditions.

The price leader, conscious that other firms will in fact follow him, will set his price as if he actually had a monopoly position. If he were not sure of the loyalty of his followers he would have to exercise caution. He could not put his price too high, lest the others raise them only slightly and thus give themselves a competitive advantage, and lest too many new firms be encouraged to enter the industry. The position of the leader is secure only when in size he dominates the other firms in the industry. There is at all times sporadic price-cutting by small competitors. And in periods of expansion it is particularly difficult for the price leader to maintain his position, because not only do the followers break away but also new firms enter the industry. Periods of contraction, however, are likely to see price

leadership at its strongest and also most welcome, for in such periods there is sentiment for price stabilization.

Although a price leader may act to bring about the same level of prices in his industry as if a monopoly prevailed, the consequences are different from monopoly. Since there are other firms in the industry the total amount of capital facilities in the industry can be larger than under monopoly. Furthermore, the profits of nearly all of the firms in the industry may be small, owing to the large investment in the industry.[1]

Basing-Point and Zone Price Systems

Price competition is eliminated or very much limited in some oligopolistic industries by means of geographic price systems, based particularly on commodities in which transportation charges are an important element in price.[2] Steel and cement are examples. Transportation charges are of peculiar importance in the sale of these commodities. The seller is interested in the net price he receives at his plant, while the buyer is concerned with the delivered price of the article. To the seller, transportation charges are costs which cannot be altered, and which are an additional burden incurred in making a sale, if indeed he bears this burden at all. In many industries (automobiles, for example) sellers do not bear the burden of transportation costs at all—they sell their products "f.o.b." ("free on board"), which means that the merchandise is placed on railroad cars or other means of transportation at the producer's plant, and that freight is paid by the buyer. There is, besides the f.o.b. system, the basing-point system of prices, which is also designed to transfer the relatively fixed costs of transportation from the seller to the buyer.

The most famous of the basing-point systems is "Pittsburgh plus." This was a price structure which existed in the steel industry more or less continuously from 1900 to 1926. Each steel company, no matter where its plant was actually situated, charged a price for steel which was the mill price *plus full freight charges from Pittsburgh.* A Chicago steel mill, selling steel to a buyer perhaps only a few hundred yards away, added freight charges from Pittsburgh to its price. The "Pittsburgh plus" system, however absurd it might seem, at least had the merits of simplicity and uniformity. It was a means of preserving price uniformity in the steel industry, where price competition had been so intense that several attempts had been made to

[1] Cf. A. R. Burns, *The Decline of Competition,* McGraw-Hill, 1936, pp. 140ff.

[2] Cf. Temporary National Economic Committee, Monograph No. 1, *Price Behavior and Business Policy,* 1940, Part II, *passim.*

stabilize prices. Pittsburgh was the basing point because it was the center of the steel industry (United States Steel Corporation's principal plants were in Pittsburgh). It was an arbitrary system, abandoned because of its doubtful legality, but it decreased price competition in an oligopolistic industry selling a product for which transportation charges are important.

"Pittsburgh plus" is an example of a single base-point system for determining prices. Multiple basing points are also used in such industries as cement, wood pulp, lead, sugar, and steel (since 1926). The multiple basing points are set at the centers of production. Each firm charges a price based on freight charges from some basing point to the buyer's locality, whether or not the selling firm is located at the basing point. Under the multiple basing-point system, (1) prices to any one buyer are identical no matter from whom he buys; and (2) buyers in widely separated localities pay different prices because different basing points are applicable to them.

Zone Price Systems

In the zone price system prices are uniform within zones; the seller either quotes delivered prices directly, or quotes f.o.b. prices and pays the full freight himself. Many goods sell at exactly the same retail prices throughout the United States, especially nationally advertised goods. Uniform delivered prices exist also for some industrial products, like aluminum, where freight is a small element in total cost to the buyer. For such products the entire country comprises one zone. Zones may be merely a part of a state, or may take in all of the territory east or west of the Rocky Mountains.

Zone systems have usually arisen as part of the growth of an industry. The zone boundaries are sometimes established to delineate sales areas "belonging" to particular sellers, but ordinarily zones are devices to keep prices in line. Generally the uniform zone prices are observed by all sellers, with the exception that the price on an individual order may be cut if the order is considered an important one. Prices differ from zone to zone. The differentials may be influenced by custom, or may reflect conditions peculiar to each zone.

Freight Equalization

Where several sellers are in competition but are situated at different points in the country, their competition can take the form of absorbing all

or part of the freight charges to the localities of buyers.[3] Suppose there are two rival producers situated in New York and Chicago who are selling identical commodities. If Philadelphia and St. Louis are two important markets, if sellers use an f.o.b. system of pricing, and if the plant prices of both are identical, the New York producer will not be able to sell in St. Louis, nor the Chicago producer in Philadelphia. At some point of time, however, one of the producers may have surplus capacity or surplus product on his hands which cannot be absorbed by the territory which he has been accustomed to serve. If under this condition the New York firm learns that a buyer in St. Louis is in the market for a large order, the New York firm may attempt to invade the St. Louis territory and to compete with the Chicago producer. One way in which the New York producer can reduce his price to the St. Louis buyer sufficiently to meet Chicago competition would be to reduce his plant price by an amount sufficient to make up for the difference in freight. Usually, however, such a step would entail so large a loss on sales in near-by areas that the gain from making occasional sales in remoter areas would be completely lost. Hence a much more expedient and likely decision would be to cut prices only on the St. Louis transaction, and by an amount about equal to the difference in freight. If prices are quoted on a delivered basis, this can be done directly; if an f.o.b. system is used, the quotation can be reduced by the discount necessary to make the sale. Some such process of "freight absorption" is likely to occur whenever geographically separated sellers of similar products compete in the same markets.

Freight equalizations, however, tend to assume conventional forms. In many industries there are systematic schemes of freight equalization more or less followed by all sellers. Such schemes are far too elaborate in detail to represent purely spontaneous developments. But the existence of schemes of freight equalization is not by itself proof of the absence of effective price competition. "All that can be said is that the more complex and conventional types of geographic price structure often spring not from natural adjustments to the market but from a conscious effort on the part of concerns in an industry to limit or eliminate price competition and that they may even be indispensable tools to this end. Thus although the more obviously artificial types of geographic price structure are frequently encountered under circumstances which give rise to suspicion of collusive restraint of trade, yet it is not necessarily true that their existence is either the direct cause or manifestation of such collusion. It is theoretically possible, for example,

[3] The following discussion is adapted from *Price Behavior and Business Policy, op. cit.*, pp. 277ff.

that aggressive price cutting should take place within the framework of such a system as well as without it." [4]

RELATIONS BETWEEN MARKETS

Thus far the processes of price formation under pure competition, monopoly, monopolistic competition, and oligopoly have been examined from the position of the seller. Throughout the *buyers* have been assumed to be actively in competition with one another. It is necessary now to drop this assumption in order to examine relations between buyers in various kinds of markets.

Pure competition *among buyers* is found in almost all retail markets. Witness the consumer. If he doubles his purchases of soap or cigarettes or anything else, the price he pays will not rise on that account. The price he pays could not rise, because the individual buys only an insignificant part of the total amount consumed. Or suppose that an individual consumer decides to pay less than the going price for some article. If he sticks to his low price offer he is able to buy nothing, because in most markets there is no individual bargaining on prices.

If an individual is the only buyer of a commodity or service, his position corresponds to that of the monopoly seller. The single buyer is also a monopolist—a monopoly buyer or *monopsonist.*[5] Just as the monopolist is able to choose from the many available the one selling price which maximizes his profits, so the monopsonist is able to choose from among many purchase prices the one which maximizes his gain as a purchaser.

Monopsony is found throughout the economy. A meat-packing company might be the only buyer of beef in a locality, an auto-assembly company the only buyer of certain auto parts, a chemical concern the only buyer of certain herbs, etc. Local monopsonies are perhaps not so common as local monopolies, but they exist in many areas, especially in the market for services. The local mill or factory is the only buyer of a certain grade of unskilled labor, the drugstore is the only buyer of the services of qualified pharmacists, the garage the only buyer of mechanics' services, etc.

The monopsonist, unlike the buyer in pure competition, within limits can choose the price he pays. If one corporation acquired all facilities for refining copper and thus became the only buyer of copper concentrates, what price would it pay? Not the lowest conceivable price, because at such a

[4] *Ibid.*, p. 278.

[5] This word is of recent coinage. It is formed on the model of the word "monopoly." Literally, a monopsonist is a sole buyer.

price the owners of copper mines would refuse to produce; nor the lowest price which would cause the copper mines to produce, because at such a price the owners of copper mines would not produce enough as the manufacturing concern could profitably use. The price paid by a monopsonist is the price which maximizes his gain as a purchaser, and which causes his suppliers to furnish the amount he wants.

Monopsonistic competition and *oligopsony* among buyers correspond to monopolistic competition and oligopoly among sellers. The distinction between the two market forms is based on numbers, but it is not so definite for buyers as for sellers. Likewise, the importance of product differentiation is not so great for buyers. In its broadest aspects, product differentiation comprises all those characteristics of a product (or seller) which are intended to lead buyers to prefer one seller to another, within the limits of substitutability between one seller for another. From the point of view of the buyers, then, product differentiation means those characteristics which lead sellers to prefer selling to one buyer rather than another, again within the limits of substituting one buyer for another. It is undoubtedly true that sellers often prefer to deal with certain buyers, usually to the latter's advantage. This applies particularly in the sale of labor services. For example, one department store in a large city may have the reputation of being a "better" store; and all of its employees shine in the reflected glory of the store's reputation. Hence this store is better able to keep a sales staff than other department stores in the city, and may even be able to pay lower wages. Thus, the sellers of labor exert a preference in seeking employment, and the buyer is able to take advantage of this in the form of a lower price.

More important is the advantage which large buyers are able to secure. Meat buyers buying from cattlemen, cigarette manufacturers buying from tobacco farmers, and automobile companies buying from parts manufacturers are examples of oligopoly buying. Such buying keeps the price low, just as under oligopoly selling the price is higher than under pure competition. Chain stores are able to buy many commodities at lower prices than the independent retailer must pay, not only because the chain store buys in large quantities, but also because the chain store uses its size as such to force lower prices (by threatening not to buy at all, it can cause reductions in the prices it pays).

Several possible buyer-seller relationships are possible. The relations on both the buying and selling sides of markets can be described as follows:

1. *Pure competition on both sides.* No one person can control the price. This condition exists in the great farm commodity markets, and was discussed in Chapter 15, pages 338 to 346.

2. *Monopoly by a seller, pure competition among buyers.* This is the classical monopoly case.[6]
3. *Monopoly by a buyer, pure competition among sellers.* This was described above, on page 371.
4. *Monopoly on both sides.* This is "bilateral monopoly," exemplified by industry-wide bargaining between a labor union and an industry. The price arrived at by the buyer and seller is "indeterminate," i.e., it is not possible rationally to say just what the price will be under such circumstances. Price is here dependent on the particular facts in each case—on whether the buyer or the seller is the better bargainer, whether the buyer could produce the product himself if he thought the seller's price too high, whether one has better knowledge of costs than the other, whether the seller is a producer also of other products, etc.
5. *Oligopoly or monopolistic competition among sellers, pure competition among buyers.* This is exemplified by the automobile industry.
6. *Oligopsony among buyers, pure competition among sellers;* example: the market for raw tobacco.
7. *Oligopoly and oligopsony.* This relationship is very common in the markets for producers' goods. For example, the relatively few railroads buy rails from the relatively few steel companies. Here also, as under bilateral monopoly, the price or prices arrived at depend largely on particular circumstances and cannot be precisely stated with general reasoning.
8. *Oligopoly (or oligopsony) on one side, monopoly (or monopsony) on the other.* Thus, there can be one buyer and several sellers, or one seller and several buyers. The monopolist (or monopsonist) is more powerfully situated than any one of the few buyers (or sellers) with whom he deals, but he is not so powerful as he would be if he dealt with many buyers (or sellers).

Each of these market forms or relationships exists in the economy. There are, furthermore, certain *relations between markets* which also determine the market structures within which prices are formed. In flowing through the different stages of production, a raw material very often passes from one type of market to another. Take a commodity the raw material for which is grown on farms. The farmers sell their product to a processing industry which consists of a few large firms, i.e., there is pure competition among the sellers and oligopoly among the buyers. The buyers have the advantage in the market, even if they do not resort to collusion in fixing

[6] See Chap. 15, pp. 346-54.

their purchase prices. The processing industry next sells its product to another industry for further processing. The second industry may consist of a large number of small firms, so that there is now oligopoly among the sellers and pure competition among the buyers, the advantage lying with the former. Thus the first processing industry has a double advantage because of its market position, which is that of an oligopoly buying from and selling to purely competitive industries.

Suppose next that the second processing industry sells the still unfinished product to another industry which prepares it finally as a consumer good, and suppose that the third industry is also an oligopoly. The second industry —a purely competitive one—is at a double disadvantage because it is squeezed between two oligopolies, buying from one and selling to another.

Now the third—an oligopolistic—industry sells the final product under brand names to retail dealers who in turn sell the item to final consumers. Among the consumers there is of course pure competition, and among the dealers both oligopoly and monopolistic competition are present. The third processing industry, like the first, has the twofold advantage of buying and selling in purely competitive markets.

RELATIONS BETWEEN PRICES OF COMMODITIES

The relations between the demands for and the supplies of pairs of commodities was discussed earlier. It was shown how the demands of two commodities are connected when the commodities are substitutes for each other or are complementary with each other. Similarly, it was pointed out how the supplies of two commodities are related, when the two are jointly produced or are rivals in production.

Consider now the price relations of commodities which are substitutes. The general principle is that *the prices of goods which are substitutes for one another move together.* If the price of grapefruit rises the price of oranges is expected to rise also, although not necessarily in the same proportion. Because as the price of grapefruit goes up the amount bought diminishes, some of the buyers switch their purchases to oranges which are a substitute product. The demand for oranges thus increases, and their price rises. Similarly, a decline in the price of a product causes a decline in the prices of its substitutes. If the price of pork declines a larger amount is bought, so that there is a shift of consumption from beef. The demand for beef is less, and consequently its price can be expected to decline.

This statement of movements in the prices of substitute goods require

some qualification. First, prices must be flexible before the changes in the price of one commodity can be reflected clearly in altered prices of another commodity. Consequently, these price relations stand out best in purely competitive markets. These relations operate most directly, in fact, in the retail markets for such foodstuffs as fresh vegetables, meat, and fruit, for these commodities differ from other retail items precisely because their prices vary from week to week, if not from day to day. Therefore, if a commodity is sold at a "rigid" price, i.e., one changed only infrequently, such as the price of a branded article, a rise in the price of the substitute good may not produce any change in price at all. Despite this possibility, however, the change in the price of the substitute causes an increase in the demand for the commodity with the controlled, or rigid, price. If the increase of demand is large enough and is expected to continue, there is occasion for the deliberate raising of the price. Thus the principle of the direct dependence of the prices between substitute commodities is effective even when prices are controlled.

The second qualification to the principle is concerned with supply. If the price of commodity A rises, how will conditions of supply affect the rise in the price of commodity B, which is a substitute for A? An increase in the price of A heightens the demand for B, but if the supply of B were perfectly elastic (the sellers are willing to supply more at no increase in price) then the price of B would not rise. Supply must not be perfectly elastic: this is the only qualification to the principle of the interdependence of the prices of substitutes.

Consider next the price relations of commodities which are complementary to each other. There are many pairs or groups of goods which are jointly demanded: Tennis balls and tennis racquets, toothpaste and toothbrushes, pens and ink, etc. If the price of one of a pair of complementary commodities goes up, what happens to the price of the other? It tends to go down. For, if the price of one goes up, less of it is bought and therefore the demand for its complementary product falls off. Thus while the prices of substitutes move up and down together, the prices of complements may move in opposite directions. In both cases, the change in one price need not be proportionate to the change in the other. Indeed, usually the changes are not proportionate.

The qualifications given for the substitute price relation apply with equal force to the complementary relation. If prices are not flexible, or if supply is perfectly elastic, the price of one complementary good may not be affected by a change in the price of the other.

The substitute relation between commodities is much more common than the complementary relation. In a sense, all commodities are substitutes for one another, because they all compete for the consumer's dollar. The complementary relation results from the combined production of goods or of their joint use in clusters of goods. Since the substitute relation dominates, it follows that prices tend to move in the same direction. The principles described for pairs of commodities hold true for several commodities and to all commodities and services. A rise in the price of one commodity is followed by increases in the prices of several commodities. The effect of the rise of one price spreads itself like a ripple through a whole series of prices, and, like a ripple, the effect diminishes the farther away it proceeds.

Price relations are in several ways conditioned by the *supply* connections of commodities. Under joint supply, which was defined previously,[7] price relations work this way: suppose that there is an increase in demand for beef, which is produced under conditions of joint supply with hides. An increase in the demand for beef raises its price. The higher price of beef brings forth a larger supply of beef and necessarily a larger volume of hides. But if there is no change in the demand for hides, the larger supply of hides must be sold at a lower price. Hence an increase in demand for one commodity can lead to a *decrease* in the price of another, when the two commodities are produced under conditions of joint supply. Suppose again that the demand for both beef and hides increases, but that the increase in demand for beef is relatively greater. It is still possible for the price of hides to fall, if their supply increases (along with the beef) more than their demand.

Just as complementarity in demand is the exception and substitution the rule, so the joint supply relation is exceptional to the general prevalence of rival supply. Commodities are rivals for resources as well as for consumers' dollars. Where two commodities are produced under conditions of rival supply, an increase in the demand for one of them tends to raise the price of the other. The reason is that an increase in the demand for one brings about a higher price for it, thus causing an increase in supply. But an increase in the supply of one necessarily diminishes the supply of the other, whose price in consequence goes up. Therefore when two commodities are rivals, either in consumption or in production, their prices move upward or downward together.

[7] See Chap. 14, pp. 332-33.

SUMMARY

The highlights of this and the preceding chapters may be reviewed at this point. It has been shown that prices perform the following functions in the present economic system:

1. Allocation of consumer goods among consumers
2. Geographical distribution of commodities among all of the centers of use
3. Adjustment of consumption to varying levels of production (as in seasonally produced commodities)
4. Guidance of the quantity and composition of goods produced
5. Determination of methods of production, subject to the limitations of technology
6. Allocation of the basic resources—land, labor, and capital—among their possible uses

The demand for any commodity or service is composed of the demands of individuals. The amount of a commodity which an individual will buy depends on his tastes, his income, the price of the commodity under consideration, and the prices of all other commodities. This can be expressed by means of the individual demand schedule. The demand for a commodity in a market is represented by a market demand schedule, which is the sum of the demand schedules of individuals who are potential buyers of the commodity.

A demand schedule shows the relation between possible prices and the amount which would be bought at each price. The relation is an inverse one: less would be bought at a high price and more would be bought at a low price.

Elasticity of demand expresses the character of the relation between price and amount bought. Demand is elastic when a change in price causes a more than proportionate change in amount bought; it is inelastic when a price change causes a less than proportionate change in amount bought. The elasticity of demand of a commodity depends upon the number and kind of substitutes which it has, the number of uses to which it can be put, and the ratio of its price to the incomes of buyers.

The income elasticity of demand for a commodity refers to variations in amount bought caused by changes in the incomes of purchasers.

The demands for different commodities are interrelated because commodities are either substitutes or complements of one another.

Supply also is represented by means of a schedule showing prices and amounts which would be offered for sale. In general, a supply schedule exhibits a direct relation between prices and amounts which would be offered for sale: the higher the price, the larger the amount which would be offered for sale. The concept of elasticity can also be applied to supply.

Three conditions of supply are to be distinguished: (1) sellers have goods on hand, (2) they are going to produce goods with existing equipment, and (3) they are going to invest in capital facilities to produce goods. These conditions are termed *market supply, short-run supply,* and *long-run supply.* Costs of production do not influence market supply, but do affect short-run and long-run supply.

In the short run, which is a period brief enough so that production facilities may be considered as given, the prices at which producers will supply are related to their costs in this way: prices must always be higher than their average variable costs, although the prices may be below full cost of production. The prices at which producers will supply commodities can also be higher than full costs of production. If the price is at a height which calls forth full production, still higher prices cannot cause any additional production; thus the short-run supply schedules of producers become perfectly inelastic at high prices.

In the long run, during which adjustments of investment in plant and equipment can be made, supply is determined by full costs of production including a normal profit. In the long run the prices at which producers will supply commodities are variable, depending upon scale of output.

Prices are determined in markets, which are areas where buyers and sellers are in touch with one another. Markets may be characterized by conditions of pure competition, monopoly (monopsony), monopolistic (monopsonistic) competition, or oligopoly (oligopsony).

Where markets are characterized by pure competition—a large number of buyers and sellers dealing in a homogeneous product—prices are determined by demand schedules and supply schedules. The equilibrium price is the price which equates the amount which would be bought with the amount which would be offered for sale.

Prices under monopoly—a single seller—are characterized by the fact that monopoly sellers themselves determine their prices within the limits imposed by the demand schedules for their products or services. The monopolist adjusts his price so that his net profits are maximized. This principle must be qualified if the monopolist does not fully know the demand for his commodity, or if because of fear of regulation or potential competition he sets his price lower than he would otherwise.

Under conditions of monopolistic competition—where sellers of branded and otherwise differentiated commodities and services are in competition with one another—prices are only one phase of competition. Prices, conditions of sale, quality of product, and advertising may all or in any combination be the terms of monopolistic competition. Where competition takes the form of price competition, prices are set by the individual seller, who is nevertheless faced with a very elastic demand for his product so that his range of choice of prices is limited.

Prices under oligopoly—few sellers—are characterized by the fact that the price charged by each of the few sellers is closely dependent upon the prices charged by the others. Thus prices can be deliberately set by oligopolistic sellers at levels identical with the height of a monopoly price. This is true particularly where one firm dominates an industry and acts as a price leader. Yet destructive price wars can always arise under oligopoly, and therefore the tendency in oligopolistic industries is for competition to take nonprice forms. Furthermore, these industries have commonly developed systems of pricing which tend to lessen the possibilities of price competition: for example, the basing-point, the zone, and the freight-equalization systems.

The function of allocating goods requires flexible prices to adjust stocks and production of goods to the changing demands of buyers. Retail prices are, in general, determined under conditions of monopolistic competition; they are fixed by the individual retailers but under conditions of very close competition. Competition in the retail field appears to be keen enough to bring about sufficient price adjustment to keep consumption and supplies in balance. Stocks of unsold goods do not pile up, except temporarily. The problem of unsold goods is generally solved by reducing prices. Likewise increases in consumer demand are nearly always met by larger supplies or by higher prices. It is only in wartime and in periods like 1945-47 that the volume of goods consumers are willing and able to buy is larger than the volume dealers have for sale.

The geographical distribution of goods and the adjustments of consumption to varying levels of production are aspects of the functioning of the price system which are performed unobtrusively.

Prices guide production—by indicating to producers what and how much to produce and to a certain extent by what means—most effectively when they are formed under conditions of pure competition. Prices under conditions of pure competition are determining, not determined. No one producer or consumer can influence these prices; rather, all must be guided by them. A high price encourages production and discourages consumption;

a low price has the opposite effect. Under conditions of monopoly, of monopolistic competition, and of oligopoly, however, individual sellers have some control over their prices. Similarly when buyers are in monopsonistic or oligopsonistic positions, they too have some control over prices. When producers themselves influence prices, the function of prices as effective guides or determinants of production is impaired but not nullified. For the power of individuals to set their own prices is more apparent than real. Under monopolistic competition and under oligopoly, the power of the individual is narrowly circumscribed by the actions and potential actions of his competitors. Even the monopolist is limited in his freedom of action by the character of the demand for his product. Thus in actual practice prices can be set by individuals only within limits: the producer is always guided by the price range within which his product can be sold.

It can be concluded, therefore, that prices function as guides to production and as determinants of methods of production even if somewhat imperfectly. Prices determined under monopolistic and imperfectly competitive conditions distort the allocation of resources to some extent; high prices keep consumption and therefore production lower than they otherwise could be.

Price is not only a determinant of the composition of the national output and the allocation of economic resources, but it also influences the level of national output. Over any period of time the size and composition of the output is the outgrowth of producers' decisions made during a previous period. These decisions are made in terms of short-run and long-run profit (price-cost-sales) expectations. The short-run expectations as shown already call for a price which will be at least above minimum average variable costs; the long-run expectations call for a price equal at least to full unit costs of production.

The prices at which the current flow of national output sells during any one day or week *in general* equal, fall short of, or exceed the expected prices which lead producers to schedule this flow of output. Current prices are among the influences which form producers' expectations of future prices. Thus if current prices and sales are not up to previous expectations, this is likely to affect adversely the short- and long-run expectations which govern the level of national output, income, and employment in the period ahead. If current prices are about what had been expected, or generally higher, the level of production in the period ahead will be about the same or may increase. Short-run expectations of price-cost-sales govern the volume of output of consumer goods and many raw materials; long-run expectations govern the production of capital facilities. These expectations of the indi-

vidual producers, relating to the prospective demand-price-cost relationships for each product or line of products, determine the level of national output in the period ahead. That is, the aggregate supply, or national output, is the sum total of individual producers' outputs; and the latter in turn depend upon the particular market relationships which have been described.

The aggregate demand for the national output is also the sum total of the demands of consumers, investors, and government. Individual demand has been analyzed on the basis of preferences, prices, and incomes. Thus aggregate demand is derived from the innumerable particular demands, just as aggregate supply is made up of particular supplies. At any given time the market demands, market supplies, and market prices determine the size and composition of the national output. And the expectations of market demands, market prices, and costs govern the future level of national output, income, and employment, as well as the composition of this output and the allocation of economic resources.

CHAPTER 17

The Prices of Productive Services

Productive services are the services of persons and property used in the production of the national output. *Productive services are those services which are demanded because they contribute to the output of commodities and services.* All productive services can be classified into the services of the four factors of production: land and other natural resources, labor, capital, and enterprise.

The necessity for separate and extended consideration of the pricing of productive services arises from the nature of modern society. In a primitive economy in which people depend upon hunting and fishing for a living, the output of services is directly connected with the wants of the people performing the services. If a person wants fish he goes fishing, and thereby performs the productive service to satisfy his want. In an expanded industrial society, on the other hand, the relationship between wants and productive services is almost always indirect. In a money economy consumer

wants are reflected through demands, i.e., how much people are willing to pay for particular commodities and services. Productive services are still another step removed from the satisfaction of wants because the demands for these services are made by producers, who will use them to produce commodities and services for consumers.

In the previous analysis of the pricing of consumer goods two basic assumptions were necessary. It was assumed, first, that people had incomes to spend, otherwise there could have been no effective demand. Second, it was assumed that the influence of factor prices on producer costs was given. But incomes and factor prices are not determined independently of each other. *Incomes, costs, and the prices of productive services are linked together and mutually determined.* Consider, for instance, one kind of productive service—labor. When wages are determined (other things being known) not only the cost of labor services to producers will be determined but also the incomes of workers.

NATURE OF DEMAND FOR PRODUCTIVE SERVICES

Demands for consumer goods[1] are demands for goods wanted for personal consumption (bread to eat, pictures to look at). Demands of this kind are not equivalent to, for example, a demand for scrap iron. Scrap iron is demanded not for the purpose of satisfying any personal want of the person making the demand; the scrap dealer does not want to wear or otherwise use the scrap iron himself. The scrap dealer wants scrap iron because he can sell it to a steelmaker. The steelmaker demands scrap iron because it is required in varying proportions to make good steel and because it usually lowers costs to mix scrap with pig iron in charging his furnaces. For each demand for productive services, such as for the services of scrap iron or of steel workers, there is no direct connection between demand and want satisfaction.

There is, however, an *indirect* connection between the demands for productive services and demands for consumer goods which is easily perceived. The steelmaker demands scrap iron, labor, and many other things to make steel because its production and sale provides him with income. The buyer of steel may be an automobile manufacturer, who requires steel and other services to make automobiles, which will finally be bought by consumers. Thus, despite the intervening factors of time and processing, productive services are demanded because people demand (or are expected to demand) consumer goods which require productive services for their

[1] See Chap. 13, pp. 292-301.

manufacture, sale, and delivery. The demands for productive services are therefore termed *derived demands,* because they are derived from the demands for consumer goods into which they enter, either directly or indirectly.

A further distinction must be made in the use of the terms *consumer goods* and *productive services.* The term *economic good* is considered to include both commodities and services, e.g., shoes and shoeshining, bread and the serving of bread at restaurants, etc. Strictly speaking, the demand for goods for personal consumption is always a demand for the services which are expected to be derived from these goods. A person does not demand bread as such, but rather the service which bread performs in satisfying his appetite; not shoes, but the services of shoes. But since the usage of the term *consumer goods* is so general it is retained here.

In the definition of "productive services," however, it is preferable for several reasons to stress the function of services. First, productive services include labor services, which even in noneconomic language are identified as "services." Second, it is preferable to stress productive services to emphasize that a thing is wanted for what it does, e.g., machinery, fixtures, store space, and other capital facilities. Third, productive services often involve commodities used in production which are indivisible or "lumpy," i.e., which exist only in certain sizes. Thus, by concentrating upon the service aspect of indivisible commodities used in production, it is possible to conceive of their services as small units which more easily lend themselves to demand and supply analysis.

The major difference between demands for consumer goods and demands for productive services is a subtle one, and the failure to perceive the distinction usually makes the man in the street a poor economist. In some of the previous discussion, the demand, supply, and price of a particular commodity were regarded in an isolated context apart from other commodities and their prices. The demand and supply schedules of other goods are affected by the price of a particular good, but these effects are generally observable only in cases of substitute and complementary goods.

When the demand for productive services of certain kinds is considered, however, it is often dangerous to ignore the relationship between one demand or price and another. People are accustomed to think of the demand for labor services in the same terms as they think of the demand for strawberries, hats, or some other familiar item. It is a serious mistake to assume, for instance, that the demand for labor services in the United States is unaffected by the prices of the services of capital, or by the prices of these same labor services, or by the prices of labor services somewhere else in the world. The demand for labor services in the United States

is not independent of these prices, because the wage rates of labor at home and abroad and the return to capital both directly affect the level of national income; the level of national income in turn affects the demand for consumer goods, which in turn affects the demand for labor services in highly important and often unpredictable ways. For example, if the prices which might be paid for labor services are important quantitatively in their effect on the national income, a demand schedule for such a service cannot be drawn up. This is true because the establishment of any price for the service will determine wages and therefore influence national income. A change in national income in turn will change the entire schedule of prices and amounts in any schedule that might be drawn up. Although this problem is not peculiar to the demand for productive services, it is far more important in this field than in the field of consumer goods.

CLASSIFICATION OF PRODUCTIVE SERVICES

The classification of productive services which is closest to the experience of businessmen is that classification based upon costs. Since costs are dependent upon the prices of productive services it is profitable to examine the type of cost classification used by most businessmen. Accountants in their day-to-day work are required to record items of cost which can be charged to the products produced in order to determine whether a business or a portion of it is operating at a profit or at a loss. In general, costs are classified according to whether or not they can be traced directly to the production of a particular commodity.[2]

Consider, for instance, the production of a type of chairs by a furniture factory producing several kinds of furniture. It is possible to determine the amount of wood, glue, screws, lacquer, and other materials required per chair, i.e., the cost of materials directly embodied in a chair. It is also possible, by observing the amount of labor directly expended in machining the parts and assembling them, to determine the cost of labor required to make a chair. There are, however, other costs of producing chairs than materials and labor directly applied to their manufacture. There is the original construction cost of the factory building and the cost of maintaining it; there is the original cost of the machines and the cost of repairing them; there is the cost of power, of foremen to supervise operations, of officers to run the company, of salesmen to obtain orders, of accountants to keep books

[2] The classification of costs presented here is different from that contained in Chap. 14 because the purpose of the analysis is different.

and estimate costs, of clerks to prepare invoices, and of stenographers to handle correspondence.

A breakdown of the costs of making chairs would look something like this:

A. Direct Costs
   1. Materials: costs of wood, glue, paints and varnishes, etc.
   2. Direct labor: wages of operators of saws, lathes, shapers, etc.

B. Indirect Costs
   1. Indirect labor: wages and salaries of foremen, maintenance men, machine setup men, etc.
   2. Factory burden: costs of power, heat, light, repair parts, depreciation, supplies, insurance, property taxes

C. Overhead (general, administrative and selling costs)
   1. Allocation of executive salaries, wages of general office labor, costs of office supplies
   2. Allocation of salesmen's salaries, advertising expense
   3. Allocation of interest, bad debts, general business taxes

In this classification the costs of making chairs are arranged primarily in terms of the ease with which the different types of productive services can be allocated to the manufacture of a particular type of chair. Costs for wood, glue, and finishing materials (A.1) can be estimated with a high degree of accuracy; labor costs for men who work directly on the specific type of chair (A.2) can also be estimated with a fair degree of accuracy under the conditions prevailing during any single period. The problem of determining the costs of supervision, repairs, and maintenance attributable to the production of a given type of chair is somewhat more difficult. A foreman may spend part of his time supervising workers making one kind of chair, part in supervising other workers, and part in filling out requisitions and maintaining records common to the production of all the chairs made in the department. This condition is true of all indirect labor costs (B.1). It is also difficult to determine just how much of the heat, light, power, supplies, insurance, taxes, etc. (B.2), should be charged against a particular commodity.

Taking the chair department as a whole, however, it is possible to determine total cost of indirect labor and factory burden on the basis of past experience. This total cost can then be allocated to each type of chair in proportion to the number produced, direct labor per chair, or some other reasonable basis. If the estimates are made carefully, for any given period

of production, the sum of the unit indirect costs will equal the total indirect costs of chair manufacture.

The costs of overhead (C) are even more difficult to trace specifically to the manufacture of a particular kind of chair. The officers of the company and their secretaries are concerned daily not only with chairs, but with all the other products made by the company. The accounting and sales staff are likewise occupied with the whole line of products sold by the company. Interest on money borrowed by the company is a general charge against the business; it is usually impossible to determine to what extent the chair department has benefited from the loan. Bad debts arise from failure of customers to pay bills for chairs and for other products, all of which may have been included in the same account. Real-estate taxes apply to the business as a whole, of which the chair department is only a part.

Productive services, therefore, can be classified by accounting into: (A) some which can be traced directly to a specific item of production; (B) some which can be traced directly only to a certain class of items; (C) some which cannot be traced beyond the business as a whole. By this kind of classification a business concern is enabled to make price-cost comparisons which are a useful guide in determining business policy. It will be noted that this classification does not make a clear-cut distinction between fixed and variable costs which was assumed for purposes of simplicity in Chapter 14.

However, this cost-accounting classification of productive services is by no means the most desirable for the purpose of analyzing the prices of productive services in the entire economy. The wages of the same man or the same material may be a direct cost in one firm and overhead cost in another. There is nothing in the cost classification which is related to the price-making forces which determine the prices of the productive services used in making chairs. In order to clarify terms it is necessary to examine the main classifications used in economic thinking of the past.

During the latter part of the eighteenth and most of the nineteenth centuries, many economists were concerned about the proportion of the national income received by the various classes in the society of their time. Being concerned with incomes of classes, they logically used a classification of productive services based upon the social classes in English society furnishing the productive service. Laborers (the "lower class") received wages for their services; landowners (the "upper class") received rent for the services provided by their land, and capitalists (the "middle class") received profits for their services and the services of their capital. A division

of the national income among workers, landowners, and capitalists seemed to correspond well with recognizable groups of income-receivers in the more industrialized countries during the period. It became customary to regard "land," "labor," and "capital" as the factors responsible for the production of all goods and services, and to consider the total product or income to be divided among the owners of these "factors of production."

In the nineteenth century the classification of the factors was revised to include not only the group of people who supply capital but also the group who undertook to bring it together with labor, land, and materials to produce goods. The *entrepreneur* (enterpriser), a person who organized and co-ordinated and directed these factors, was distinguished as a fourth factor. Since new goods were constantly being discovered, produced, and offered for sale, and since the structure and level of demands were always changing, the entrepreneur emerged during the nineteenth century as a kind of hero in the play of economic forces. It was he who foresaw the opportunity to develop or expand markets, calculated the risks, and plunged in to capitalize on the possibility of turning an honest dollar.[3] If all went well, he could pay off his creditors and make a sizable "profit" for his trouble; if his calculations were faulty, he paid off his labor, lost his shirt to his creditors, and received nothing or went into debt for his time and trouble. Thus, the classification of "factors of production" was expanded to include, in addition to land, labor, and capital, the entrepreneur, whose income consisted of "profits."

To this day it has been customary to follow, if with many refinements, the classification of productive factors or economic resources which had its origin in the eighteenth and nineteenth centuries. "Land" has come to be recognized as a leading species of a large genus–factors in fixed supply; and "capital" is no longer furnished solely by "capitalists" as the term is now understood. Even the term *entrepreneurship* was detached from a one-man "entrepreneur" when it became difficult to identify the person who could be called *the* entrepreneur in such corporate giants as the United States Steel Corporation and General Motors; the emphasis is now placed upon a function rather than a person. However, it is still true that the term used to describe each of the factors of production encompasses a class of productive services subject to about the same price-making forces. Each factor is distinguished from the other because the forces which determine its price are different from the forces determining the prices of the other factors.

Thus, returning to the classification of services based upon business costs,

[3] See Chap. 7, pp. 166-72, for a further discussion of the role of enterprise.

it is obvious that the business-cost classification can be broken down into these four factors. Materials can be reduced, by following them back to their own sources of supply, to a combination of some or all of the four factors. The same breakdown can be applied to every item which is purchased by the firm, such as insurance, power, supplies, and advertising. Even tax payments can be broken down into a payment for labor, capital, and land made by the government. Depreciation and interest represent costs of capital, and bad debts represent a negative profit, or loss, to the entrepreneur. The services of land, labor, capital, and entrepreneurship (or enterprise) are therefore the four classes of productive services which are used in price analysis in this book.

RELATION OF PRICES OF PRODUCTIVE SERVICES TO INCOME

Consider the relationship between the prices of productive services and the incomes of owners of productive services. The prices of labor services may be expressed as certain amounts per hour; these rates multiplied by the number of hours worked at each rate during the year will equal the total annual income from labor services for an individual worker. The sum of the labor income of every individual performing labor services will obviously be the total income from labor services for the year. The same process can be applied to interest, rent, and profit.[4]

If the payments made by any business unit are examined, it is found that certain payments are made directly for productive services and certain are for materials and services furnished by other business units. It is obvious that the direct payments for productive services are counted in reckoning up the total national income, but what of the payments made to other firms? If these payments are traced to other firms, it is found that part of them are paid out directly for productive services and part are paid for materials and services supplied by still other firms, and so on, ad infinitum. Eventually, however, all payments will resolve themselves into payments for productive services of the four basic types. The sum of the payments for productive services is the national income.

The relationships discussed can be illustrated by Chart 16.

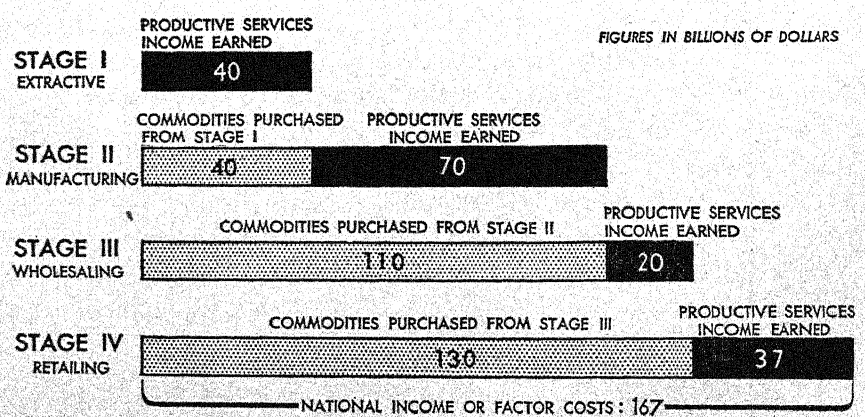
The four bars labeled Stage I, II, III, and IV represent levels of production, which are conceived to encompass extractive (agriculture, mining, forestry, etc.), manufacturing, wholesale, and retail types of business. The length of the bars represents the factor costs of product at each stage, the

[4] The way in which profit may be treated as a price is discussed in Chap. 21.

cost (value) being made up of both payments to factors for productive services and payments to producers in a preceding stage. Included in payments to the factors is the income credited to the accounts of owners of the factors (though not necessarily paid to the owners). Beginning with Stage I, which is assumed to consist simply of combining productive services to produce products, payments equal to 40 billion dollars are made to suppliers of productive services or to their accounts, the same amount being realized from the sale of the products to the next succeeding stage. Stage II adds 70 billion dollars more in payments to productive services, and sells products for 110 billion dollars to Stage III. Stage III adds 20 billion dollars in payments to productive services, selling the output for 130 billion dollars. In Stage IV, the products finally reach finished form and are sold to consumers. After the addition of 37 billion dollars more in payments for productive services, the total value of final output (net) is 167 billion dollars. From this analysis the following statements may be made:

1. Payments for productive services total 167 billion dollars (40 plus 70, plus 20, plus 37). This is obviously the income of society, since all other payments—totaling 280 billion dollars—are simply transfers from one stage to another. The 280 billion-dollar figure plus the 167 billion dollars of income payments represent the total value of transactions in the economy.
2. At each stage factor costs are covered, since each stage receives an amount at least equal to its payments for productive services and payments to other stages.

Chart 16 PRODUCTIVE SERVICES AND NATIONAL INCOME



The sum of the payments for productive services is the national income; as shown in Chapter 4, these incomes or factor costs are the principal costs of producing gross national output. When indirect business taxes and depreciation costs are added to factor costs the total equals gross national output.

DEFINITIONS OF FACTORS OF PRODUCTION AND THEIR PRICES

The description of sources of productive services in terms of labor, land, capital, and enterprise has been adhered to by many generations of economists because it contains the elements of a classification useful for determining the prices of productive services. If land, labor, capital, and enterprise are considered in terms of their *supply characteristics*, rather than in terms of the social classes with which they are or were associated, this classification of productive services is useful for price analysis.

Land, the principal natural resource, and rent were the chief subjects of concern to economists of the first half of the nineteenth century, particularly in England. The condition of the working people at this time was generally close to subsistence. It was obvious that while population could increase rapidly, the amount of land was physically limited and could not be increased (or decreased) to any appreciable extent. Land yielded very large returns to landowners because of its scarcity.

Obviously, many resources used in production have supply characteristics like land. For instance, during World War II automobiles had the supply characteristics of land: the supply could not be increased. When demand rose, as it did most markedly, the price of used cars skyrocketed without bringing forth new automobiles. Similarly, services of the most popular movie stars at any time are strictly limited and the price paid for their services is enormous. The impression should not arise, however, that services characterized by fixed supply always command high prices. During any short period of time, a fixed supply may also result in extremely depressed prices. When, as was true for many years, so much textile-mill capacity existed that a large portion of it was not needed to supply the market, the prices of textiles were kept at generally unprofitable levels because many producers were willing to sell textiles at prices which allowed no return at all on their equipment. The same condition prevailed in many industries in the United States and in other countries after the depression beginning in 1929.

It is necessary to define land in terms of its supply characteristics for the problem under consideration. Thus, land must be defined as a natural resource which is, *for a period under consideration*, fixed in supply. The

price of its services is called rent. However, when other resources are in a state of fixed supply, their prices are called quasi-rents, to indicate that the price analysis applicable to land also applies to them. Plants, machinery, and even certain labor services are in fixed supply under certain conditions; the analysis of the pricing of their services, therefore, parallels the analysis of rent. On the other hand, natural resources under certain conditions must be regarded as if they were capital because their supply is subject to change. In the long run, land is the only factor in fixed supply; in the short run, many factors may be in fixed supply.

The services of persons must be regarded in a separate category because human beings in a civilized society are by law free and cannot be owned.[5] The right of free persons to withhold their services and to enter the occupations of their choosing, and the inability of any person to impose involuntary servitude upon another, distinguish the services of human beings from the services of any other factor of production; thus, a separate analysis of labor services and the payment for them, principally wages, is required. Labor services are services of human beings performed to obtain a stated reward which consists of something besides the performing of the services itself. Or, as some writers state it, *labor is human effort exerted for gain.* The stated rewards, or prices, for which services are rendered are wages and salaries. At this point, all salaries and wages are grouped into one class: wages.

Mere expenditure of effort is not "labor." Many people make great exertions in playing golf for pleasure, in working for the success of a church bazaar, or in trying to get elected to a nonsalaried political office. Such activity is not labor in the economic sense of the term because it is not performed for a stated reward. Labor services are services offered for a price; if the service commands no price, it is outside the realm of this discussion (except insofar as it indirectly affects prices for labor services, as when amateur entertainers throw professionals out of work). If all or part of the reward for the service is contingent upon the success of the undertaking, i.e., if it is in effect a profit-sharing arrangement without a stated reward, then all or part of the service ought properly to be considered under enterprise. The common-sense meaning of the term *labor* will suffice in economic discussion.

The term *capital* has a great variety of meanings, virtually all of which are correct when considered in proper context. One reads of the "capital"

[5] If slavery were the accepted practice, labor services might be considered in the short run to be similar to the services of land, and in the long run to be similar to the services of capital.

of a firm, of "capital stock" of a corporation, of the struggle between labor and "capital," of the necessity for "capital" to start a business, of borrowed "capital" and of "capital" derived from reinvested earnings, of fixed and circulating "capital," of money and real "capital."

Perhaps the most widely held notion of capital is that it is money. By this test, countries which had the greatest amount of money outstanding during the wartime inflation, such as China, should have had the greatest capital. In such countries, however, the amount of capital facilities was very small. Money in itself should not be considered as capital.

The concern here is with capital as a source of productive services, and with the price paid for such services. Machinery, plants, and other types of capital facilities discussed earlier, and stocks of materials and finished goods are required to carry on most businesses. *The services of goods which are, for the problem under consideration, durable but exhaustible are the services of capital goods.* The services of capital (goods) are distinguished from the services of land by the fact that the amount of capital services is exhaustible (and presumably must be constantly replenished) whereas the amount of land is fixed. The following discussion will be concerned with the effects of the prices of capital services upon their supply.

Capital goods consist of plants, machinery, buildings, improvements, tools, and stocks of finished and unfinished goods.[6] Such goods yield their services only over a period of time; it requires time not only to derive all the service from buildings, machinery, and the like, but also to manufacture raw materials or to sell stocks of finished goods. Since in a modern economy money must be invested in capital goods, it is necessary that loans be advanced to obtain their services. Loans are usually advanced in terms of money, or they can be valued in terms of money. *The price paid for loans* (and thus for the services of capital goods) *is interest.* A complication arising from the fact that almost all money loans bear interest, whether they are devoted to production or consumption, will be discussed later.

Although the price paid for capital services is interest, the total return paid to capital is not. The essence of capital is that it is exhaustible; therefore, if the owner of capital is not to become gradually impoverished he must "maintain" his capital. If, for instance, a man invests $100,000 in capital facilities and expects to maintain his holdings over the years, the business must return to him not only an annual interest on his investment but also an earned amount over and above this annual interest to repay the amount he has invested; quite apart from "profit," the business must

[6] See the partial list on pp. 70-71.

earn enough to return his $100,000 before the income-earning power of the business is exhausted. It is for this reason that depreciation charges are made against income (though they are not always earned), and a sinking fund or other reserve is set up against long-term borrowing.[7] Similarly, the fertility of the soil is a part of the farmer's capital, and also in a sense part of the nation's capital; if it is not conserved both farmers and the nation as a whole may become poorer.

The nature of enterprise or entrepreneurship has already been indicated. *Entrepreneurial services consist of directing and co-ordinating other productive services for the purpose of realizing a profit—a return over and above the costs of all the other productive services and the depreciation of capital.* Like labor services, entrepreneurial services are supplied by human beings who are themselves free. Unlike labor services, these services are undertaken without promise of definite payment, and unlike the provision of labor services, the provision of entrepreneurial services may actually yield a negative return, i.e., entrepreneurs may actually be money out of pocket after expending effort.

Entrepreneurs can exist only in a dynamic economy, one in which unexpected changes occur. If economic life were static, and economic activity consisted of repeating without change established processes, the function of co-ordinating productive services would be ingrained in regulatory institutions and in the individual, social, and political habits of the people. Economic life would be reduced to routine, and there would be no occasion for any producer to adapt himself to change or to seize upon a new opportunity. In an economy of unpredictable change, characterized by the development of new products, new techniques, and new markets, the services of enterprise perform the tasks of directing other productive services into new lines of activity. For performing this function successfully, entrepreneurs earn profits, and for failing to perform it successfully, they suffer losses.

The price of entrepreneurial services is called profit; unlike all other prices, this price can be either positive or negative. The supply of entrepreneurial services, it will be found, is markedly influenced by the level of profits, which acts not only as a governor for the supply of such services but also, in a competitive economy, as a governor for the prices of goods. If profits in any line are high, the amount of entrepreneurial service de-

[7] This point has been elaborated in connection with the annual production of capital facilities, i.e., the need to replace capital used up in the production of national output. See Chaps. 4 and 5.

voted to that line is likely to increase, and when it does, the supply of goods in that line is likely to increase with a consequent reduction in prices.

To classify productive services into four main classes—the services of land, labor, capital, and entrepreneurs—is to give them their widest interpretation in relation to their common characteristics. Obviously, all land is not identical, nor is all labor or capital, and probably no two entrepreneurs are ever exactly alike. The problem of pricing productive services is often that of pricing many types of services which fall within one of the four general classes. Because the sources of productive services are numerous and varied, the term *agents of production* is used to indicate the multiplicity of specialized sources. Once the principles of pricing the four major classes are understood, they can be applied easily to the problems of any sub-classification. Thus, once it is known how the services of land are priced, the same principles can be applied to rural land and urban land, to farms in Iowa and to building lots in New York City.

THE MARGINAL PRODUCTIVITY THEORY

It has already been noted that many generations of economists were concerned with the prices of productive services because of their interest in the share of the income which was distributed to broad classes of people —workers, landowners, and capitalists. The distribution of income among persons depends not only upon the prices of productive services but also upon the ownership of sources of productive services by private persons. A wage earner may be rich, for instance, because he owns capital as well as because he has control over his own services. The distribution of income among *persons*, therefore, is a subject which can be considered only after the pricing of productive services has been determined. For this reason it will be discussed later.[8]

To understand what forces determine the rate of pay of particular types of labor in particular areas and the earnings of capital in specific industries in an area, one would have to examine all the circumstances which might affect the prices of such productive services. Nevertheless, certain principles can be applied which at least explain the nature of these forces. Fundamentally, these principles are simply an elaboration and refinement of the supply-demand analysis already given in previous chapters.

The theory which is commonly used to obtain a general explanation of the pricing of productive services is called the marginal productivity theory. Although, like all theories, this one is valid only if certain simplifying

[8] See Chaps. 37-40.

conditions are assumed, it nevertheless illustrates the operation of powerful forces at work in the present-day economic system. To avoid confusion and particularly to avoid misapplications of the theory (some of which will be illustrated later) the simplifying conditions should be kept in mind. Without such simplifying conditions, which are like the rigid controls applied in laboratory experiments, the discussion would be extremely complicated and the reader would become lost in a welter of exceptions and qualifications.

1. *It is assumed that entrepreneurs seek always and in every way to maximize profit.*
2. *It is assumed that technology remains unchanged.*
3. *It is assumed that the sources of productive services are limited and constant in quantity.*
4. *It is assumed that the tastes and incomes of people are constant.*
5. *It is assumed that pure competition prevails in the market for both consumer goods and productive services.*

   It is further assumed that people know where they can buy or sell goods at the best prices, and all buyers and sellers of productive services know the prices paid for such services and each seeks the most advantageous price. There is complete mobility of both goods and sources of productive services.

Under these conditions, the demand for productive services by a given firm may be analyzed. It will do little violence to the facts to consider that technology is such that it is possible to produce a given product by means of various combinations of productive services. The product may, for example, be produced with much labor and little capital, or little labor and much capital. Airport building by thousands of peasant farmers using hand tools in China, and the same construction by use of modern machinery in the United States, represent widely varying combinations of labor and machinery, although the differences in cost may be small. What combination of productive services is the most economical?

A relatively simple type of business, a parking lot for cars in a city, will illustrate how a combination is selected. The product is parking space for automobiles; the parking rates are given and the rate of pay of helpers is also given. The parking-lot operator must determine how many man-hours per week he should hire to conduct his business; i.e., how much labor to use in combination with his land and his own services as entrepreneur.

By attempting to run the business alone, the operator is faced by several contingencies if the lot is to be open from 8 A.M. to 12 P.M. on weekdays—

or 96 hours per week—and if he himself does not care to work more than 72 hours. Some people park for nothing when he is absent; many park haphazardly because of lack of direction and therefore waste space; some people leave their cars locked and so block the exit of other cars; finally, the lot will gain a reputation for being unsafe for lack of constant supervision. Realizing this, the operator considers hiring attendants, who are assumed to be available for as little as 8 hours' work per week.

If he adds 8 hours of help per week to his own 72 hours, he can collect from patrons faster during the rush period and direct some of the parking at that time. He still cannot keep the lot fully supervised, however, and will still lose some collections during the unsupervised period. If he hires a second 8 hours of help, he can institute better control over parking during the busy period and lessen the unsupervised period, thereby increasing both his volume of business and his collections. A third 8 hours of help either would permit supervision of the lot at all times or would allow still better direction of parking during the rush period: either improvement will add to his receipts. A fourth 8 hours of help would clearly permit continuous supervision, and would permit two or three men to direct parking during the busiest periods. By progressive increases in the amount of help employed both supervision and direction of parking would be improved, so that an eighth 8-hours unit, for example, would permit such added service as delivering cars to the driveway for people. At some point, however, the addition of extra help will add nothing more to the business receipts and, because of confusion and inattention among the excess number of employees, the business may eventually suffer. Suppose this condition results after the ninth 8 hours of help is added.

It can be seen in Table 43 that the increase of business receipts derived from adding help is greater per unit for the first few 8-hour periods than for the later periods.

The addition to the total product resulting from the addition of a unit of labor service is termed the *marginal product* of labor. The marginal product is measured in physical terms, not in value. *The marginal physical product multiplied by the price of the product is called the value of the marginal product.* The *value of marginal product* decreases as more help is hired for the reasons which have been described. Eventually, when additional help can add no more service which will increase the volume of business, the value of the marginal product falls to nothing (at the hiring of the ninth 8 hours of help). At an indeterminate point beyond this, (X 8 hours of help), the total receipts of the business will decrease because there is an excess of employment.

Table 4.3 RETURNS FROM HIRING HELP AT A PARKING LOT

1. Units of work	2. Number of 8-hr. units worked	3. Total weekly receipts	4. Addition to total receipts (value of marginal product)	5. Receipts per 8 hrs. per man (value of average product)
Owner works alone 72 hrs. per week	9	$120	—	$13.33
With 1st 8 hrs. of help	10	170	$50	17.00
With 2nd 8 hrs. of help	11	210	40	19.09
With 3rd 8 hrs. of help	12	245	35	20.42
With 4th 8 hrs. of help	13	275	30	21.15
With 5th 8 hrs. of help	14	300	25	21.43
With 6th 8 hrs. of help	15	310	10	20.67
With 7th 8 hrs. of help	16	316	6	19.75
With 8th 8 hrs. of help	17	320	4	18.82
With 9th 8 hrs. of help	18	320	0	17.78
With Xth hrs. of help	—	315	—5	—

The *value of the average product* is found by dividing total receipts by the number of units of labor service. It rises with the addition of more labor until the marginal product is less than the average product.

This example illustrates a force which is always at work in all forms of production: the *law of diminishing returns.* It may be stated generally as follows: *when additional amounts of one factor (or agent) of production are used in combination with a fixed amount of the other factors of production, the marginal and average products of the factor (or agent) being added to the combination will eventually fall.* In the example, labor was added to a fixed amount of land and entrepreneurship. The marginal product, it will be noted, falls at once, although close examination of the example will show that during some part of the range from one to nine 8-hour units worked the marginal product must have risen. The average product in the example first rises and then, after 14 units of work per week, falls. The point at which the average product begins to fall is the *point of diminishing average returns.* The point at which the marginal product begins to fall is the *point of diminishing marginal product or return.* Although the law of diminishing returns relates to the decline of physical returns or product as units of productive service increase, it can be illustrated in terms of value (as in the example).

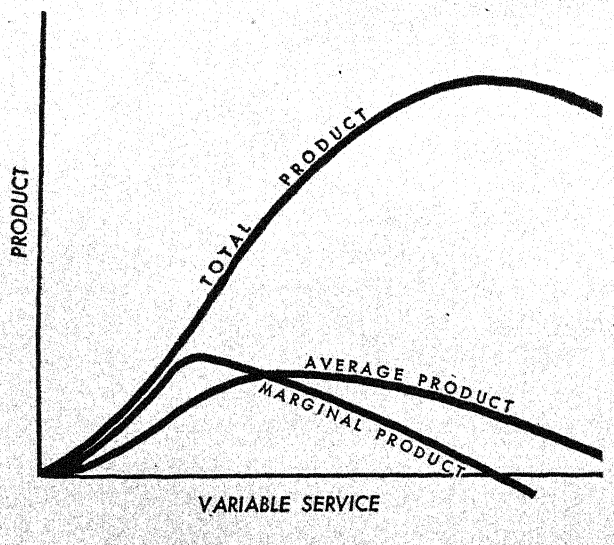
The law of diminishing returns can be illustrated geometrically, as in Chart 17. The vertical axis measures the product in physical units resulting from the use of varying quantities of one type of productive service (meas-

ured on the horizontal axis) with a fixed quantity of all other services. The product may be imagined to be parking space for automobiles in a parking lot of such size that increases in the product and of the variable service are infinitesimally small; thus changes in amount of product can be represented by smooth curves. Product in this example is measured in value terms.

The total-product curve shows the amount of the total product produced by varying the amounts of one type of productive service used with a fixed amount of the other services. The average-product curve is derived by dividing the total product by the number of units of variable service, e.g., it shows the output per man. The marginal-product curve shows the rate of change in the total-product curve, i.e., the change in the total product per infinitesimal change in the variable service used.

So long as the total product increases at an accelerated rate, the marginal-product curve rises. When the total-product curve rises at a progressively slower pace, this means that the marginal-product curve is falling. When the total-product curve begins to fall, the marginal product becomes negative. The entrepreneur will never operate beyond this point, because to add variable service at this point, even if it were free, would reduce his total product. Neither would he limit his operations at the point of maximum marginal product (under pure competition), because the return per unit of variable service can be increased beyond this. The entrepreneur will add labor up to the point at which the value of the marginal product is about

Chart 17 **DIMINISHING RETURNS**



equal to the wage. This is nearly always past the highest point on the marginal product curve, but short of the point of zero marginal product.

It may be noted that the average- and marginal-product curves have similar shape and bear the same relationship to each other as the average and marginal *cost* curves in Chart 12. There is, of course, a mathematical relationship between product and cost. If average product per unit of variable service is high, average cost must be low, and vice versa. In fact, the average variable cost is simply the average product divided into the price per unit of the variable service. The marginal cost is simply the marginal product divided into the price per unit of variable service. For example, if the price of the variable service is $5 and the marginal product is 10 units, the marginal cost is 50 cents; if the product is 5 units, the marginal cost is $1, and so on. In the subsequent discussion the significant concept is the marginal products of productive services and the tendency for the marginal products eventually to diminish as units of productive services are increased.

Demand for Productive Services

A basis for the demand for productive services has now been established. Beyond the point of diminishing marginal returns, a producer would be willing to pay more per unit for small quantities of productive services than for large, because the smaller quantity is worth more per unit to him. In the parking lot illustration, since the first eight hours of help add $50 per week to receipts, it would pay to hire a person if he will work eight hours for less than $50. If parking lot help can be hired for $5, it would then pay to hire up to eight eight-hour units of help, since the seventh unit hired adds $6 to total receipts and costs only $5. The operator is indifferent toward employing the eighth unit of help because it would add only $5 to receipts while the added cost remains at $5.

The schedule showing the value of the marginal product is in fact a demand schedule of the individual firm for labor service, showing that more would be hired at lower than at higher wage rates so far as the firm is concerned. The reason for the willingness to hire more labor at lower than at higher wages is to be found, then, in the operation of the law of diminishing returns.

Obviously, there is some combination of labor with the other co-operating factors which is most advantageous to the firm. This combination under pure competition is beyond the point of diminishing returns for additions of labor to the combination, because any firm would employ additional

labor at a constant additional expense per unit to itself if in so doing it could constantly increase the marginal physical product and, therefore (remembering that the price of the product is constant), the value of the marginal product. What, in general terms, is the most economical combination of productive services?

In employing productive services, the firm adds units of service of any given type to the combination so long as it pays to do so; so long as the marginal product resulting from the addition of labor, for example, sells for more than the cost of the additional labor, it pays to add labor to the combination because by doing so total cost is increased less than total revenue. Diminishing returns being in operation, the marginal product of labor declines as more and more units are added to the combination until it equals the price of the productive service, at which point it is no longer profitable to add additional labor. All competitive firms, under the conditions assumed here, follow the rule of equating the value of the marginal product of a productive service with its price because it pays to do so.[9]

It was assumed that a fixed quantity of the other types of productive services already existed in the combination (a given amount of land and enterprise in the parking lot). It might be possible, however, to modify the proportion of all factors used in a combination. To a certain extent, labor can be substituted for land by increasing the amount of work applied per unit of land. Labor can be substituted for enterprise by spreading the entrepreneur's services more thinly. The entrepreneur of a parking lot may undertake to set up another lot, for example.

Productive services are not indefinitely substitutable for each other, because as one type of service is substituted for another, or increased in relation to the others, the law of diminishing returns sets in. There is, therefore, one combination of the different types of productive services which is the *optimum*, or *least cost, combination.* Suppose that a unit of labor service costs \$1 and a unit of capital service costs \$2, and that at one stage in the combination the *marginal products of the two are equal.* Clearly, this would not be the optimum combination; it will pay to reduce the amount of capital services and increase the amount of labor services in the

[9] The principle that the value of the marginal product equals the price of a unit of productive service is to be understood in terms of infinitesimally small additions of the variable service. If these two quantities were not equal, the entrepreneur would sacrifice an infinitesimally small addition to the difference between his costs and receipts. In the same sense that $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8}$, etc., may be said to equal 2, so additions to the variable service may be said to be made until the value of the marginal product equals the price of the variable service.

combination. As this is done, the marginal product of labor services falls and the marginal product of capital services rises through the operation of the law of diminishing returns. When the marginal product of capital has risen and the marginal product of labor has fallen to a point where the latter is one half of the former, it no longer pays to substitute labor for capital. If the marginal product of capital at this point is 10 units of product and the marginal product of labor is 5 units of product, further substitution of labor for capital will lower the marginal product of labor possibly to 4, and raise the marginal product of capital possibly to 11. Spending $2 more on labor will add something less than 8 units to the product; the same amount spent on capital will add 11 units to the product.

The optimum combination of productive services is reached when the ratio of the marginal product of one productive service to another is equal to the ratio of the prices of the productive services. This rule is general and applies in all cases. In the present example, the ratio of the prices of productive services is $1:$2. When the ratio of the marginal products is 1:1, it pays to substitute labor for capital, thereby raising the marginal product of capital and lowering that of labor. When the marginal product of labor becomes 5 units and that of capital becomes 10 units, a ratio of 1:2, then the optimum combination is reached, since further substitution results in a ratio of 4:11, or 1:2¾.

The limited substitutability of one kind of productive service for another serves to limit the size of a firm. If all productive services in the combination could be increased proportionately without increasing the unit cost of production, there would be no limit upon the size of the firm producing a particular product. The size of a cotton mill, for instance, might be anything from a one-loom unit to a million-loom one. The powerful force which eliminates this possibility in most industries is that services of at least one type cannot be expanded proportionately together with all the others without loss in efficiency. This type of service is that of control, or enterprise. As all other services are increased, the problem of controlling them becomes more and more difficult, with the result that a point of the most economical combination of productive service relative to entrepreneurship is reached.[10]

Prices of Productive Services

Each firm under pure competition employs various types of productive services in such a way as to equate the value of the marginal product with

[10] See Chap. 14, pp. 324-25.

the price of each type of service. When this result is achieved, each firm will be in equilibrium so far as its employment is concerned. There is, however, a general condition necessary for the equilibrium price of each type of productive service, the price which was assumed in the foregoing examples. *The price of the productive service must be such that it equates the total supply and total demand for that productive service.* If this condition is not established, obviously the price of the productive service will either rise or fall, and the value of the marginal product of the productive service to every firm will no longer be equal to the price of the service.

As an illustration, assume that the price of a particular type of labor service is established at which the total supply exceeds the total demand. Granting the other assumptions regarding competition, knowledge, mobility, etc., the following may result: (1) the owners of the unemployed services will almost certainly undercut the market, and the purchasers will aid and abet the process; (2) to the extent that technological processes permit, firms will not only substitute this type of labor services for other services, but also employ additional labor services in conjunction with their present combination; (3) the prices of services most competitive with the lower priced labor services may in turn be reduced, but the prices of services required for co-operation with additional labor services in production will be increased; (4) the prices of some products requiring considerable amounts of the cheaper services may be reduced, and the prices of other products requiring considerable amounts of the co-operating services may be raised. Thus equilibrium is achieved only when it no longer pays to substitute one type of service for another. Substitution will no longer pay when the price of each productive service—which equates total demand and total supply—is equal to the value of its marginal product.

Therefore, the price of a productive service depends upon the total supply and total demand of the services, the total demand being determined ultimately by the value of the marginal product of the service to all of the individual firms employing it. The equilibrium price of a productive service is one at which (1) the value of the marginal product to each firm is equal to the price of the productive service, and (2) the price of the productive service is such that demand for it is equal to its supply.

Limitations of the Marginal Productivity Theory

A serious criticism of the marginal productivity theory is that it assumes that the combination of the productive services can be varied, when in fact technological conditions often prevent variation. In employing men to dig

a ditch, for instance, labor must be employed in constant combination with capital—one man, one pick; one man, one shovel. Giving a man an additional pick or shovel will add nothing to the total product; putting two men to work with but one pick to use will add virtually nothing to the total product. In this criticism it is likely that too narrow a view is taken. For a given product such as a ditch, a very considerable variation in the combination of the factors is possible, given enough time to change the combination. To increase the proportion of labor, lighter and cheaper picks could be used, even to the extent of using sharpened sticks; to increase the proportion of capital in the combination, pneumatic drills or even mechanical ditch-diggers could be used. Uneconomical hand-labor methods were used on many work relief projects for the unemployed during the 1930's to increase the use of labor, in much the same manner (though not for the same reasons or under the same conditions) that predominantly hand methods are used in some Asiatic countries. Indeed, under the conditions prevailing in such countries, what are considered "backward" methods are probably the most economical.

For those cases in which technological processes definitely fix the proportions of services which must be used, it is still possible to apply the marginal productivity analysis. Services which must be used in fixed combination can be priced as though they were but a single service, and the analysis may be carried on in those terms.

Another criticism is that the marginal productivity theory is vitiated or made unrealistic by the existence of monopoly and by the absence of mobility and knowledge. This type of criticism is aimed not so much at the theory itself as at its usefulness. There is no doubt that under conditions of monopolistic competition, the buyer of productive services does not determine the value of additional services by multiplying the marginal physical product by the price of the product. Since additional output is likely to reduce the unit price of the entire product, the value of additional services is determined by multiplying the marginal physical product by the revenue or receipts added to total receipts by the sale of the marginal physical product. Since the *marginal revenue* is less than the price of the product of a monopolistic firm, the firm will not employ a productive service costing $5 per unit up to the point at which the marginal product sells at $5. Instead, the firm will employ a productive service up to the point at which the marginal product multiplied by the marginal revenue equals the price of the productive service. If the buyer of productive services is engaged in monopsony or monopsonistic competition, there will be a similar result. The marginal productivity theory does not explain prices of productive

services in the absence of pure competition, but it provides a useful starting point for the analysis of such prices under other conditions.

The values of the marginal products of productive services may differ from their prices when the owners of productive services cannot, from lack of knowledge or mobility, supply the services where the value of the marginal product is highest. A man may work for 40 cents an hour as a common laborer when he might be employed for 75 cents an hour on a semiskilled job which is available but not known to him. The interest on mortgage loans in Alaska might be 10 per cent but lenders elsewhere may be ignorant of this, or may be unable to obtain sufficient knowledge to gauge the risks properly. In these cases, the theory indicates merely the probable direction of change in the prices of productive services.

Another criticism of the marginal productivity theory is that it does not correspond to the practices of businessmen. It is argued that businessmen who have never heard of the "value of marginal product," and who could not calculate it if they knew what it was, are nevertheless quite successful. This criticism is of little consequence. Although businessmen may not use the same term to describe the process of calculation which they perform, they do in fact make calculations. When they say, "It just would not pay us to hire more men," they have in mind considerations such as those discussed. Many of the elements stressed by the theory, such as a variable combination of productive services, are obscured in ordinary business practice by technological considerations which limit the possibility of varying the combination. Machines are designed and factories laid out to produce the product with a certain combination of men and machines and factory space. A closer approximation to the conditions assumed by the theory is found in the planning of a new plant, because in the planning various combinations of services must be considered. The plan actually adopted is usually thought to be the least cost combination for the type of production contemplated, and it is usually only one of several that will have been considered. Although the marginal productivity theory assumes for the purposes of analysis a more precise calculation than is ordinarily possible in practice, to say that businessmen do not act in the manner presupposed by the theory is to say that they are not interested in knowing whether it pays to take on or lay off workers, or whether it pays to consider different techniques of production using various combinations of productive services —that they are not interested in maximizing profits.

As for the general applicability of the marginal productivity theory, in almost every particular situation to which it may be applied it calls atten-

tion to the basic forces at work and the probable directions of change in the economy. Without the use of the marginal productivity theory it would be difficult to offer a rational explanation of such matters as the differences in wages in different countries or different occupations, or the prevalence of hand labor in low-wage countries, or the existence of tall buildings in cities, or the fact that low-wage countries do not depress the standard of living in high-wage countries by oversupplying them with cheap goods.

CHAPTER 18

Wages

The prices paid for the use of personal services in the production of the national output are commonly termed *wages*. For purposes of the present analysis, wages include salaries, commissions, fees, as well as wages in the usual meaning of the term—hourly, daily, or weekly rates of pay. When the word "wage" is used to mean a price, it is always the amount paid per unit of service, per piece, per hour, per day, etc. The number of units of service performed by a person during a year, multiplied by the appropriate rates, determines the annual wage income of the person. Incomes derived from "wages and salaries" comprise between 65 and 70 per cent of national income.

Wages in a broad sense here include some income which is ordinarily not regarded as a wage or salary payment. For example, an independent storekeeper, or farmer, may not, as the economist does, regard all or part of his income as a wage or salary; his income is often called profit. If the storekeeper hired himself out as a manager of a store, or if the farmer were employed by someone else as a farm manager, their incomes would then obviously be in the wage category. When working on their own account they earn "wages of management," which is an amount equal to their maximum incomes had they been employed in similar capacities elsewhere. This is termed *imputed* or *implicit* wages, in contrast to explicit, or direct, wages. Although implicit wages are not represented in the usual statistics

on wages, because they cannot be quantitatively estimated, they form a type of income important in economic analysis.

Another distinction of considerable importance for some purposes is that between money wages and real wages, which parallels the distinction made earlier between money income and real income. *Money wages are simply given amounts of money paid per hour, day, week, or on some other basis for the use of personal services. Real wages are the purchasing power of this amount of money.* If prices of all goods bought with wages increase 10 per cent and money wage rates rise likewise, real wages remain unchanged. A change in money wage rates may be accompanied by no change in real wages, and vice versa.

Although the Clayton Act (1914) solemnly declares that "labor of a human being is not a commodity or article of commerce," it must be treated as such in economics, since the general principles which determine the prices of the services of inanimate things apply equally to the services of labor. In treating labor services as "articles of commerce," it is not intended that the dignity of the individual be impugned, but only that the forces which determine wages be made clear.

EARLIER THEORIES OF WAGES

Wages, like many other subjects in economics, are still often regarded in the light of old theories which have been in part or whole discounted. These theories are nevertheless significant for they are, as it were, the intellectual baggage inherited mostly from the nineteenth century. They make possible an understanding of the social attitudes of the past, some of which persist to this day. Particularly, they explain the legislation of the nineteenth-century reform movements as related to wage questions.

Malthus's famed subsistence theory is implicit in his *Essay on Population*, published in 1798. Malthus among others held that the level of wages is generally determined by the cost of subsistence, an idea which is today recognized in the belief that wages are based on the "standard of living." According to the subsistence theory, the level of wages is governed by the tendency of population to increase at a greater rate than the means for supporting it. Population has been known to double within a period of 25 years, but its means of support—chiefly the produce of the land—are subject to the law of diminishing returns and cannot keep pace with an expanding population. A given increase in the population cannot, therefore, result in a proportionate increase in the food supply, and as population increases the standard of life must decline, with the eventual result that

population is forced to diminish. Population diminishes through a rising death rate caused by malnutrition, by a declining birth rate from delayed marriages or some kind of birth control, and by wars directly or indirectly induced by pressure of population. In this way the level of wages is held by force of circumstance at approximately the level necessary to provide subsistence for workers and their families. A higher level would lead to an increase in the population which would in turn cause wages to fall and begin the process of diminishing population until the subsistence level was again reached. And a lower level of wages would reduce the population, or slow down the rate of increase, causing wages to rise (because of labor scarcity) again to the subsistence level.

Although admittedly there are parts of the world where the subsistence theory of wages provides a reasonable explanation of the facts—chiefly the Far East—historical experience proves that it is especially inadequate when applied to areas where industrialization in some degree is found. Even where the subsistence theory does apply, it can be considered to be only a special case of the marginal productivity theory, i.e., the value of the marginal product yields no more than a subsistence wage.

A major reason for the failure of the subsistence theory to explain wages in the well-developed economic systems is its exaggeration of the practical consequences of the law of diminishing returns. Although the law of diminishing returns is of almost universal validity when the proportions of one type of productive service in a combination are increased, its effect can be overcome by improvements in technology. Thus means have been found to increase enormously the yield of crops from the land by improved cultivation, use of fertilizers and soil-building crops, crop rotation, better strains of plants, and the like. Moreover, the subsistence theory assumes that the supply of arable land cannot be increased markedly, when actually the opening up of new lands for settlement as well as improvements in transportation in the form of railroads and steamships has greatly increased the supply of land which could be economically utilized.

The subsistence theory also slights the effect of a rising standard of living upon social behavior, i.e., in changing the relation of population to its means of support. As the standard of living increases people become accustomed to a higher standard of consumption and are reluctant to abandon it. Individuals delay in marrying, and as the age of marriage rises the birth rate necessarily falls; or again, parents delay the marriages of their children—either to increase the parents' income or to prepare the children with education—with the same effect on population rates. In urban areas particularly, women exploit their opportunities to work and therefore post-

pone marriage in favor of careers. Finally, the desire to maintain a high plane of living and to provide adequately for their children leads many people to limit the size of their families. All these influences affect the purely biological forces which in the subsistence theory are largely assumed to operate without external change.

A second theory of wages, advanced by many economists in the first half of the nineteenth century, the so-called wages-fund theory, emphasized the *demand* for labor rather than its supply (population). The wages-fund theory rightly reiterated the fact that production takes time. A considerable amount of labor must be expended—and paid for—over a period of time before a finished product can be produced. A fund of capital must therefore be provided in order to pay workers during the period required for production. This fund can be thought of in terms of a sum of money, although actually it is a store of commodities available for purchase by the workers. The wages-fund theory maintained, therefore, that wages were limited at any one time by a wages fund (stock of commodities). Should one group of workers obtain higher wages the wages of other workers would be thereby reduced. Should all workers attempt to obtain increased wages they would inevitably be unsuccessful, for this would eventually reduce the fund available because the return on capital would be less in the future. It can be seen that the wages-fund theory was well adapted to serve the interests of businessmen during an expanding period of capitalism when the demand for capital was great and when wage increases were considered harmful.

The wages-fund theory is valid to the extent that the level of wages does depend upon the supply of other resources and that investment gives rise to employment. There is, however, no fixed stock of commodities predestined to provide the real wages of labor, just as there is no given amount of money set aside for this purpose. It is true that during a period of maximum production of capital goods or war goods there will be a limited supply of consumer goods, but even under such conditions the distribution of the available supply between wage earners and the rest of the population is not fixed. Under certain conditions, moreover, if savings tend to become excessive, increased wages may actually maintain or increase employment by altering the disposition of income between saving and spending, i.e., by increasing the propensity to consume. More important at this point, however, is the fact that labor productivity may increase in ordinary times as a result of the very process of investment on the part of those who advance capital for employing labor, so that a higher level of wages is maintained out of increased output.

A theory of wages which has many supporters today is the *bargaining* theory. In this theory wages are held to depend upon the relative strengths of the parties to the wage contract—the employer and the employee—in relation to the level of demand and supply of labor. When a powerful union goes on strike for an increase in wages during a period of high employment, employers often cannot resist. Under depressed conditions workers are frequently forced to take wage cuts because they fear that business firms might otherwise reduce or cease operations. The bargaining theory emphasizes that instead of a single "correct" wage there is a range of *possible* wages, any one of which may be established by bargaining. The upper limit of the range *at any one time and place* is determined by how high a wage the employer can pay without closing down his business, the lower limit by how little the worker will accept and still stay on the job. The upper limit may be above the value of the marginal product of labor because entrepreneur, management, and capital may be highly specialized and forced to accept very low incomes. The lower limit may be much below the value of labor's marginal product because labor may be in a similar position.

There is no doubt that the bargaining theory helps to explain how wage levels are established in particular situations, especially under conditions of monopolistic competition. It must be remembered that the marginal productivity theory is but a general guide under freely competitive conditions. The bargaining theory is not general enough, however, to explain what wage level is consistent with full employment (a wage equal to the value of the marginal product when the supply of labor equals the demand). And it would be difficult indeed by using this theory to determine what level of wages ought to be paid to workers in a socialist state in which all the means of production are in the hands of the state. (If a socialist state wishes to make the best use of its labor, it will be guided by the marginal productivity theory.) The fundamental objection to the bargaining theory is that it mistakes form for substance. It shows the process of bargaining over wages but it does not distinguish a bargain as desirable or not, nor does it explain whether the wage arrived at can be maintained or whether it will cause unemployment. For answers to these questions, the marginal productivity theory provides considerable guidance.

THE DEMAND FOR LABOR

Labor is not homogeneous. Instead of laborers, there are stenographers, clerks, machinists, and a great variety of specialists, as well as unskilled

laborers. The demand for labor under actual business conditions is a demand for specific types of productive services rather than for labor services in general. However, labor services are here taken to be homogeneous for purposes of analysis of levels in general. The principles based upon this assumption are found to be equally appropriate to each type of labor services for which separate demand and supply schedules can be conceived to exist.

It has been shown that the price of any productive service in equilibrium equates its demand, which rests upon its productivity, and its supply. If one considers the demand for goods to be given, the state of technology to be constant, and assumes that long-run conditions apply, several principles which will influence the demand for labor and the level of wages may be stated.

1. The demand and consequently the price for labor tend to be higher when the supply of the other productive services is greater. In the preceding chapter it was seen that as services of one type of productive agent are substituted for services of other types in a productive combination, the marginal product of the service being increased will decline. The converse of this principle is that the marginal product of the service which is proportionately reduced in the combination will rise. For example, when land is so plentiful that it is free, the use of labor past the point where the return per unit of labor is at a maximum is not profitable because a less than maximum return may be avoided by using more land. And with capital, the greater its amount, the smaller the proportion of labor that need be employed with capital to give employment to all workers at a given level of wages. These principles help to explain why the wages of labor in the United States are generally higher than elsewhere in the world. The higher level of wages is not the result to a significant extent of any mental, moral, or physical superiority of the American worker; rather, the amounts of natural resources and of capital facilities in this country are larger relative to the amount of labor than in other countries.
2. The demand and consequently the price for labor tend to be higher as labor is used more productively. As technology improves, i.e., as the same amount of services can be used to obtain a greater output, the whole level of the marginal productivity schedule is raised. Although this principle can be illustrated only by taking particular

cases, as when a cotton gin is substituted for the hand cleaning of cotton, it can be readily accepted for labor services in general.

3. When the price of labor is low, other things remaining unchanged, there is greater employment of labor. This principle is simply a declaration that labor services are subject to the law of diminishing returns. However, it obviously cannot be applied to the problem of unemployment during depressions without drastic qualifications.

The three principles set forth above apply to the demand for labor for the economy as a whole. The total demand for labor is, of course, made up of the demands for labor by individual firms. For each firm, a demand schedule for labor may be conceived which relates the amounts which would be employed to the various possible wage-prices of labor. A demand schedule of a firm for labor under conditions of pure competition can be illustrated by the hypothetical example in Table 44.

Table 44 **DEMAND SCHEDULE OF A FIRM FOR LABOR**

Number of units of labor employed	*Marginal physical product of labor*	*Price of the product*	*Value of the marginal product (wage)*
1	20 units	$0.20	$4.00
2	18	.20	3.60
3	16	.20	3.20
4	14	.20	2.80
5	12	.20	2.40
6	10	.20	2.00
7	8	.20	1.60
8	6	.20	1.20
9	4	.20	.80
10	2	.20	.40

The amount of labor employed depends upon the prevailing wage, and the value of the marginal product of labor, which is the marginal physical product multiplied by the price of the product. If one unit of labor is employed and adds 20 units (each selling for 20 cents) to the total product, then the value of the marginal product of labor is $4. If the wage is less than $4, it is profitable to hire the unit of labor. As more units of labor are employed, the marginal physical product falls because of the operation of the law of diminishing returns, and the value of the marginal product also falls. If a unit of labor cannot be hired for less than $1.20, 8 units will be used. The schedule of the values of the marginal product

in Table 44 is the individual firm's demand schedule for labor. Since the demand for labor by the industry of which a firm is a member is a composite one it exhibits the same behavior as the demand schedule illustrated here.

This demand schedule for labor of a firm is based on the assumption that the price of the product remains unchanged as more labor units are employed. This assumption is proper under conditions of pure competition. However, if all firms in the purely competitive industry add labor there is likely to be an increase in output large enough to influence the price of the product. Thus in calculating the demand for labor by an industry both factors must be considered: first, a decline in the value of the marginal product because of the law of diminishing returns; and second, a decline in the price of the product because of expanded output. It will be recalled that under pure competition the firm has no effect upon its own prices, but the industry as a whole does. Taking this possibility into account, the demand schedule for labor of an industry, like that of the firm, shows that larger amounts of labor are employed at lower wages, but the industry demand is probably less elastic than the demand by the firm.

A demand schedule for labor for the economy as a whole is difficult to conceive. Such a demand schedule would be based on the assumption that it is possible to estimate the amounts of labor that would be employed at various levels of wages at a given level of national income. But this is invalid because various levels of wages call forth various levels of national income, since the level of wages is a major component of the level of income. It would serve no useful purpose, for example, to speculate on the effect of a universal cut in wages of 10 per cent upon employment, assuming national income, hours of work, and other factors to be constant, when by cutting wages by 10 per cent, national income at the same moment would be reduced by an average of about 7 per cent (since wages are about two thirds of the national income). In considering the demand for labor by individual firms or industries, it is safe to assume that the national income remains constant because it is little affected by a single firm or industry. The analysis of the effect of a general wage cut or wage increase is by no means a simple one, for the reason that other variables do not remain constant when the level of wages changes. A different kind of analysis from that applicable to firms or industries is needed for such cases.

Several principles of some practical use regarding the demand for labor by particular industries may be deduced from the foregoing.

1. The demand schedule for labor becomes higher as the product produced becomes more valuable. It has been shown that the amount of labor demanded by a firm (or industry) depends upon the value of the marginal product under competitive conditions. This value is simply the marginal product multiplied by the price of the product. A higher price for the product arising from increased demand for the product (not as a consequence of monopoly or minimum-price regulation) will result, therefore, in a higher demand schedule for labor by the industry; and more labor will be employed at the former level of wages, other things remaining the same.
2. The more elastic the demand for the product, the more elastic is the demand for labor making the product. If, for example, a 5 per cent cut in wages results in lower prices, the more elastic the demand for the product the greater is the proportionate increase in sales and, therefore, the increase in the employment of labor to make the product. To take an extreme case, suppose that the elasticity of demand for the product is zero, i.e., the amount demanded is the same whether the price is a little higher or a little lower. A reduction in wages resulting in lower prices will not, then, increase the employment of labor, nor will an increase in wages resulting in higher prices decrease the employment of labor. Yet, given time enough, an increase or decrease in wages will tend to influence the employment of labor because of the eventual substitution of one resource for another, including labor.
3. The smaller the proportion which the cost of labor bears to the total cost of the product, the more inelastic is the demand for labor. If air-line ticket agents took a cut in pay of 50 per cent, and the cut were passed on in lower passenger fares and freight rates, the effect on the amount of air transportation demanded and probably on the number of agents employed would be imperceptible. If, however, all air-line workers took a wage cut of 50 per cent and rates and fares were reduced correspondingly, the effect of the reduction might increase the amount of air-line transportation used and the number of air-line workers generally.
4. The better and cheaper are the substitutes for labor of a given type in making a product, the more likely is the demand for that type of labor to be elastic. If cigars may be made either by hand or by machine, an increase in cigar makers' wages will lead to the introduction of machinery and a reduction in employment, especially if the machinery is more efficient than human labor. On the other hand, where machinery cannot be substituted for labor, as in watch repairing, an

increase or decrease in wages will have little effect upon the amount of employment in the industry.

THE SUPPLY OF LABOR

The supply of labor is dependent upon (1) the size of the population, (2) the number of people willing to work, (3) the number of hours that people are willing to work, and (4) the amount and the quality of work performed during working hours. The influence of wages upon the size and composition of the population is so remote and indirect in modern countries that it is advisable to consider the study of population a separate subject and for purposes of discussion to assume that its size is given.

The proportion of the total population in the labor force is influenced by the social and legislative strictures against the labor of women and children. It is also influenced by the standard of living and the extent to which more than one breadwinner in a family is necessary to maintain it. Under certain circumstances the size of the labor force can be increased independently of changes in population. During World War II patriotic motives and high wages induced an increase of 9 million persons in the labor force (including the armed services) between May, 1940, and May, 1944. This increase was 6 million greater than would have normally occurred, and resulted from the postponement of retirement by overage workers, the shortening of the schooling period, and the entrance into the labor force of women who would not ordinarily seek jobs.

The supply of labor, in terms of man-hours, can be increased or reduced simply by lengthening or shortening the hours of work of a labor force constant in size. During periods of high employment, the labor supply is increased by lengthening hours; conversely, when there is much unemployment, shortened hours are demanded by unions to spread the work. Regulation of maximum hours of work and imposition of overtime rates are motivated partly by this same idea.

The willingness of the individual worker to supply more effort in a given period of time is in part dependent upon the wage he is paid. The worker weighs the advantage of the wages he would earn from working harder against the effort required by doing so. It is likely that at some wage level the worker will prefer an easier pace to more income, and will supply no more labor, and perhaps even less, at higher wage rates. In general, however, higher wages will induce potential workers into the labor force: housewives to work part time or full time, older workers to delay retirement or retired

workers to return to work, young people to forsake continued education for immediate positions.

The higher wages in a specific industry may, of course, draw labor from other industries, so that although the total labor supply is only slightly increased (or not at all) one section of the supply may be greatly expanded. Usually this occurs when workers are drawn from closely related occupations, as during World War II, when farmers, maintenance workers, independent odd-jobs men, and others who had some experience with tools were drawn into the building trades by the tremendous expansion of facilities construction.

Wages and the Marginal Productivity Theory

Assuming for purposes of illustration that all labor is of the same kind, under long-run freely competitive conditions there is an equilibrium wage at which the demand for labor is equal to the supply. This wage equals the value of the marginal product of labor in all its uses. If the supply of labor is increased by immigration or some such measure, other conditions remaining the same, under long-run conditions a new equilibrium wage would eventually be established at a lower level. The employment of more labor, when existing supplies of the other productive services are given and when technology remains unchanged, can be accomplished by increasing the proportion of labor to the other productive services in all lines of production. But increasing the proportion of labor in the combination of productive services lowers the marginal product of labor. Therefore, the additional labor can be employed only at a lower level of real wages than existed before. Under the same assumptions, if the supply of labor is reduced the marginal product of labor in all its uses would be raised and a higher equilibrium wage would result.

The process by which either a higher or a lower equilibrium wage is achieved is complicated, since prices of some goods and productive services are raised and others are lowered. In actual practice, of course, the change from one equilibrium position to another cannot be observed because the increase or decrease in the supply of one productive service never occurs alone; other unrelated changes during the same period of time obscure the process. The effect of an increase in the supply of labor alone is obscured and counteracted by changes in the technique of production, in the supply of capital, in the level of prices because of monetary influences, in imports and exports, and other factors. It must be remembered

too that the effect of increasing or decreasing the supply of labor described here is the long-run effect under competitive conditions. In the short run under monopolistic competition there is at best only a tendency toward the long-run effects.

Since the analysis has been concerned with labor as a total and undifferentiated supply, its usefulness at this stage is limited to cases affecting most or all workers. The restriction of immigration by the United States since 1920 is premised (on the basis of economics, at least) on the validity of the principle that an increased supply of labor tends to lower the level of wages. Again, there is today considerable effort to maintain or increase the national supply of capital facilities because it is known that the marginal productivity of labor is thereby increased. These economic policies and many others justify giving some attention to hypothetical analysis. The same type of analysis also helps to explain differences in wages, which is considered next.

WAGE DIFFERENCES

Within a given locality different occupations pay different rates, and even within an occupation workers may be paid unequally. There are also wage differences between localities, both on a general level and for the same type of job.

The problem of wage differences is illuminated by considering whether wage differences would exist in the hypothetical condition called the long run under pure competition. Under such conditions, there are no obstacles to mobility of the factors of production which are moveable and everyone has full knowledge of the market. It may be assumed that tastes of the people will be similar to what they are now, and that production of a variety of economic goods similar to the present variety is carried on.

Under such conditions, satisfaction of demand requires labor services of many types: unskilled, skilled, manual, and mental. There are among people wide differences in innate ability or natural aptitude, both physical and mental, for particular types of work; and there are varying degrees of general education and specific craft training, as well as direct experience and job qualifications (or lack of them).

It can easily be seen that there would be differences in wages even in this ideal type of economic system. The number of people of relatively limited ability would be large, and they would, because of their numbers, be forced to accept wages equal to the values of their relatively low marginal products. The number of people who can master all the skills required for engineer-

ing, accounting, medicine, or management of enterprises would be limited, and their marginal products and consequently their earnings would be high. Moreover, within each occupation there would be differential earnings related to the differences in productivity of individuals.

It is also likely that there will be geographical differences in earnings for the same degree of skill in the same occupation. Because of differences of climate, soil, topography, rainfall, and similar attributes of nature, the cost of food and housing varies among geographical areas. An area with warm winters and good soil is characterized by lower costs of food and housing than an area with cold winters and poor soil, where houses must be more substantially built, food and fuel imported, and warmer clothing purchased. Thus the value of the marginal product and the wage of a worker may be less in low-cost-of-living areas than in high-cost-of-living areas without inducing him to shift his residence. Even in two areas with the same cost of living, workers may prefer one to another because of other advantages not reflected in the wage, such as a more desirable climate, proximity to good vacation country, and the like.

But beyond differences in the training and ability of workers and between geographical areas, it is obvious that the conditions of competition, knowledge of the market, and mobility of workers in the world are not ideal in the sense just described, and are therefore causes of still further inequalities in wage levels and rates. Differences in wages between occupations are due not only to differences in ability; skilled or professional jobs usually require long periods of training which many people of the requisite ability cannot afford. To become a doctor or lawyer one must normally spend from six to nine years in college and professional school, the cost of which is prohibitive to most people, despite the advantages of state education and subsidization through scholarships. The transfer of adult workers from one type of work to another requiring a long period of training is likewise limited by the cost of supporting a family during the interim and the cost of changing residence. The entrance of workers into skilled jobs is similarly limited by the cost of training, since such jobs usually require some period of apprenticeship at relatively low wages. Entrance into unskilled occupations requires little preliminary training, and the peak income of the job is reached shortly. Therefore, people who cannot afford to undertake a period of training with low income or at their own expense are forced into unskilled jobs, and the number of them exceeds that which would be expected on the basis of varying abilities among workers. The value of the marginal product of labor in such jobs is consequently low. Another influence upon the supply of workers for particular occupations is the presence or absence

of opportunity. The potential supply for higher paid jobs is greater than the actual supply, because some low-paid workers are ignorant of better job opportunities or of how to obtain better jobs even if they know of their availability.

Occupational wage differences are further accentuated by restrictions imposed by the concerted action of trade unions and professional associations. Unions may limit membership and thereby exclude potential workers from the occupation, or they may discourage newcomers by establishing high initiation fees or lengthy apprenticeship periods. Seniority rules may be established which limit promotions to higher paying jobs. Associations of professional people often employ similar if seemingly more dignified tactics. Bar association examinations for lawyers can be made progressively more difficult, to keep down membership in the profession; and entrance requirements into qualified professional schools can be raised to exclude people from training, or even the number of "qualified" professional schools can be limited.

Occupational wage differences are also influenced strongly by the powers and attitudes of employers, both of which may depend on the extent to which they can control the instruments of government. Employers in certain areas not only have opposed unionization of their employees, but have also opposed provision for adequate relief for the unemployed either in doles or in free transportation to transfer workers to other areas.

The advantages and disadvantages of jobs are often weighed by workers on bases other than wages. "White-collar workers" usually earn less than manual workers because the social esteem attached to clerical and office jobs is held to compensate for less money. Some firms may obtain workers of a given type for lower money wages because of better working conditions and better personnel management. The supply of workers to a large and well-established firm may be maintained at lower wages because of a greater likelihood of steady employment.

Geographical differences in wages result not only from the differences in the value of the marginal product in different areas but from many other causes as well. Within a country, they depend in part upon custom and the lack of both occupational and space mobility of workers. Industries have moved into rural regions where low wages of agriculture are typical because the paid wage of the industry is supplemented by workers' ownership of land. Their products from the land raise their real incomes and the industry is therefore able to obtain labor at a cheaper rate than in a metropolitan area. On the other hand, mobility is often difficult for a worker. He must move himself and his family some distance to obtain a higher wage. He

must calculate whether the cost of moving (including temporary unemployment) will be compensated by increased wages over whatever he considers his probable period of employment. Nonwage factors—home ownership, social affinities, schooling, etc.—are important too. When the disparity of wages is found between two countries the probability of movement is further affected by differences in language, customs, and more importantly, by legal barriers to immigration.

Differences in wages between men and women in the same occupation could exist under long-run competitive conditions only if the marginal productivity of women differed from that of men. Since women are in the main less adapted to extreme physical exertion, occupations involving such effort pay higher wages to men than women (if they do not exclude women entirely). Under certain circumstances, men and women in the same occupation represent two markets for labor because demands and supplies are different. The supply of female labor is affected by the fact that the alternative opportunities for women are more limited than those for men. The "opportunity cost," i.e., the cost of wages needed to meet the competition of other available jobs, of employment for women workers is likely to be lower than for men workers. Presumably, if the value of the marginal product of the male workers is not higher than that for the women workers, all men will eventually be driven out of occupations for which women are most suited.

Male workers may actually have a higher productivity over a period of time than women workers in the same occupation under certain conditions. If the occupation requires some period of training on the job at a cost to the employer (because of the cost of extra supervision or failure to adjust the entrance rate of pay to the productivity of the new worker), then men are preferred because they are generally more permanent as employees than are women. Men are also less subject to absenteeism than women and are probably worth slightly higher wages on that account. Better working conditions must ordinarily be supplied to women than men, and this may be partially compensated for by the difference in wages, as may added costs imposed by protective legislation for women with which employers must comply. But the main factor remains the lack of alternative opportunities for women and their weak bargaining power.

WAGES AND INVENTION

Technology has long been considered by workers as a potential threat to their employment: machines are regarded often as direct causes of unem-

ployment. In the beginning of the Industrial Age in England, workers rioted against and attempted to destroy the new laborsaving machines in the textile industry. A similar attitude is exhibited today by many groups of organized labor who impose union rules against the use of machinery or improved techniques, and by some who require that human labor be paid in idleness if a machine is used (some musicians, for example). But these attitudes are not realistic for the economy as a whole, as is apparent when technology and labor are examined in a context of long-run competitive conditions.

It may be assumed that if an invention is put into use in an industry that enables a firm to maintain its production with half as many workers, labor costs are therefore reduced and prices are (under competitive conditions) lowered. Now assume that the elasticity of demand for the product of the firm is zero. Then the same volume of goods will be sold at the new lower price as before, and consumers will save through lower prices an amount equal to the wages of the displaced workers. Demand for other products will be increased by the amount saved by consumers and reduced by the amount of wages no longer paid to the displaced workers. This condition cannot persist, however, because the unemployed workers will offer to work for less. The wages of competitive types of work will fall and the unemployed will be absorbed as other industries substitute labor for other services at the lower wage. A new long-run equilibrium will be established at a lower wage for the type of labor displaced by the invention. It would appear, therefore, that the invention has resulted in a reduction of wages, but this need not be so. Although the level of money wages may be reduced, the purchasing power of wages has been raised by the reduction in prices brought about by the introduction of the invention. The total economy has gained because for the same amount of effort more goods are now produced and the standard of living is necessarily higher than before. The real income of labor will be increased by the change if the invention adds to the existing supply of capital, as is likely, because an increase in the supply of capital will raise the productivity of labor.

Whether labor in general benefits more in the long run by the reduction in prices than by the reduction in money wages depends upon the extent to which wages must fall to absorb the displaced workers. Generally, the result of the introduction of new machinery has two effects: first, it tends to displace workers for whom the machine is substituted, and second, it sets up a complementary demand for workers to make, maintain, and repair the new machines. The net displacement of workers is, therefore, never as large as it appears to be. Furthermore, if the elasticity of demand for the

product is higher than zero, and this will nearly always be true, some and perhaps all of the displaced workers might be rehired in the same industry. If this happens the adjustment will be diffused among other industries which experience some decline in demand because of the shift in the structure of the total demand.

Another result of the introduction of an invention may be a redistribution of labor between occupations with a consequent effect upon occupational wage differences. The invention may displace skilled workers in favor of unskilled, thereby increasing the wages of the latter and reducing those of the former.

The short-run effects of inventions may differ considerably from the long-run effects stated above, particularly when inventions are introduced under monopolistic conditions. The benefits of invention result primarily from the reduction in prices achieved through their use. Monopolistic producers who do not pass on the savings in costs may thereby retard the adjustments necessary for the re-employment of displaced workers. Organized workers may also resist wage cuts necessary for the process of reabsorption and thereby delay the transition. These effects indicate that if a competitive economy is to make the most economical use of resources relative to consumer demand in a dynamic economy, flexibility of wages and prices is necessary.

It is likely that if a series of inventions should occur within a short time in predominantly monopolistic industries, serious unemployment would result. The series of inventions under monopolistic conditions might raise profits because price reductions are not made on the products. If the profit receivers choose to increase their savings rather than to spend their additional income, the volume of investment must increase or the additional savings resulting from monopoly profits will generate a reduction in income and employment.

EFFECTS UPON WAGES OF UNIONS AND MINIMUM-WAGE LAWS

It has been argued long and often that "artificial" restrictions—governmental or otherwise—on the operation of a free-enterprise economy cannot improve real wages. Wages under the marginal productivity theory would be equal to the value of the marginal product of labor services, and an increase in wages could not be maintained unless the value of the marginal product rises. To force an increase in wages under such circumstances would simply create unemployment. This view is extreme because it lifts the marginal productivity theory from its setting (see the requisite assump-

tions on page 395) and applies it without modification to conditions existing at any given time. In the actual operation of the economy, there are several ways in which wage increases obtained by union activity or by legislation can be maintained without causing unemployment in the short run.

In the first place, many employers are not faced with a fixed supply price for labor. The wages which many firms must pay depend in part upon the employment by an individual firm. If a single firm is a large employer in its area, a reduction in its employment will tend to lower wages, and an increase in its employment will tend to raise wages. The firm cannot therefore consider the supply of labor to be infinitely elastic at a particular wage; rather, it must consider what effect its own employment will have on the wage. If it increases employment the firm must consider the increase in wages of both the additional workers hired and all workers of a similar type already employed. Thus the firm will hire workers up to the point at which the additional cost of labor from hiring additional workers is equal to the addition to its revenue from increased production. When it hires more workers, the wages of all its workers will increase and the addition to its costs from hiring more workers is much greater than the additional wages paid to the newly employed workers. Thus, with any given demand for labor, the firm will increase its employment less if wages rise with increased employment than if wages did not rise as it hires more men.

A wage fixed either by union agreement or by law at any level up to the marginal cost of hiring additional workers will not cause less employment but more. An employer may have employed 100 men at a wage of $1 per hour. He may be willing to hire 10 additional workers because the demand for his product has increased and each of the 10 additional workers will add, say, $1.10 per hour to the total product. To hire 10 additional workers, however, he may have to increase the wages of all his men by 2 cents an hour. The additional cost of hiring 10 more workers will therefore be $2.00 per hour more than the hourly rate of $1.02 of the 10 men, because 100 men must be raised 2 cents per hour if the 10 additional men are to be hired at $1.02. On the other hand, if a wage of $1.09 per hour were established as a minimum wage (i.e., for all workers), he would actually find it desirable to add 10 workers, since for every worker added he would add only $1.09 to his cost and $1.10 to his revenue. His whole business would be less profitable, but his employment would be greater.

If the monopsonistic conditions described are generally prevalent in the economy, general wage increases resulting from union activity or from the passage of laws could raise employment as well as wages in existing

enterprises. However, new enterprises may be discouraged and expansion of old enterprises reconsidered, and these actions may serve to offset the favorable effects of the wage increases on employment under monopsonistic conditions.

In the second place, the value of labor's marginal product has at times increased without any increase in wages. If the general level of prices has increased, for example, the same marginal product of labor will sell for a higher price. Since price increases occur gradually, profits usually rise at first without bringing a corresponding rise in wages. Union activity can secure for workers these wage increases more quickly than ordinary competitive forces, and since the value of their marginal product has increased, union demands will not reduce employment.

This would seem to imply that wage adjustments ought to be made with changes in the cost of living of workers. Wage adjustments based upon a cost-of-living index would conform to changes in the value of the marginal product of labor services if the cost-of-living increase reflects a more or less uniform and general increase of all prices. However, if there is a cost of living increase simply because of a program of government restrictions on agricultural output, for example, then there would not necessarily be a change in the value of the marginal product of labor in industry; there would be no basis in the fact of a cost-of-living increase for a general increase in industrial wages unless industrial workers in considerable numbers left industry for farming.

Another development which is most likely is that the cost of living will gradually fall as improvements are made in the processes of production. Such improvements serve to raise the marginal product of labor, and one way in which this increase in productivity can be passed on to workers is by maintaining money wages despite the fall in the cost of living. It is apparent from the discussion of these various possibilities that the linking of wages with the cost of living should not be automatic and where wages are so adjusted, adjustments should be tempered with consideration of the reasons for the change in the cost of living.

In the third place, the efficiency of labor may be associated with the level of wages. As wages are increased, efficiency may improve because of better health, better morale, or better living conditions for workers. This result is the more likely when the previous level of wages was relatively low, and is often the unexpected result of a minimum-wage law or union activity which raises the wages of very low-paid workers.

Wage increases obtained by union action or by legislative measures may cause disruptions and perhaps unemployment in a firm because the employer

is inefficient. Because of monopsonistic competition in the labor market some employers may have persisted in business only because they paid their workers less than the going wage. Raising these depressed wages to the level paid by the more efficient employers may require the employer to improve his efficiency or abandon operation. On the other hand, the attempt to increase wages in firms which are already highly efficient and which are already paying wages equal to the value of the marginal product of workers will probably reduce the firm's demand for workers unless its prices are raised and the demand for its product is highly inelastic. At any one time, many firms may be temporarily in this position because of their monopolistic advantage in the market for their product; furthermore, increased wages and attendant increased prices may encourage the development of substitutes which over a period of time will result in a reduction in the firm's demand and consequently in its employment. In any event, if the demand for the product of the depressed-wage firm or industry is elastic at the time of the wage increase, employment will be reduced almost at once.

If highly competitive firms in a certain area are all paying low wages and none is very profitable (e.g., parts of the textile and clothing industries), then the forced establishment of a higher wage will almost certainly reduce employment. Depressed-wage areas often exist because a local surplus of labor has developed from the displacement of some industry, or from an influx of immigrants of limited mobility into the area. Unemployment may well be caused by the higher wage, and workers will be induced to move with or without governmental assistance into other areas where other or similar industries can afford to pay the minimum wage. Establishment of the minimum wage serves to concentrate the public's attention on such areas and may actually hasten the process of adjustment which was bound to come anyway in the long run.

GENERAL WAGE CUTS AND UNEMPLOYMENT

It is commonly argued that since price reductions generally increase sales, the proper solution for unemployment during depressions is a general wage reduction. The validity of this argument can be questioned effectively.

In the first place, the economic logic of this proposal is weak at many points. A reduction in price of a particular commodity is usually followed by an increase in its sales. Such a price reduction could come about because of a fall in the wages of workers employed in making the commodity. A

reduction in wages in this instance probably would not affect the incomes of the people buying the product. This special case, however, leads to error when it is generalized. A general reduction in wages cannot be considered without taking into account the effect of such action upon the income payments of the economy. A general reduction in wages would reduce the demand for goods through its effect upon wage-earner incomes. Under freely competitive conditions, the prices of goods would fall to the same extent that wages are cut, and if nothing else happened the beneficial effects upon employment of the wage reduction must be the cut in prices which follows. If the argument for a general wage reduction is a good one, the argument for general price reductions to increase employment by increasing sales should be equally good—or bad.

In the second place, the argument for a general wage reduction is usually made in the context of a business depression, when unemployment results from a general decline in aggregate demand. Under such a decline few of the requisite assumptions for the validity of the marginal productivity theory will in fact be present. Particularly, the extent to which labor can be substituted for capital is limited at such a time because the processes of production cannot be revised immediately. The existing capital equipment of the country is geared to the employment of labor in relatively fixed proportions with capital: a wage reduction alone cannot quickly bring about a general change in the combination of productive services. Even those producers about to install new equipment would hesitate to adopt fewer laborsaving methods unless they knew that the wage reduction would last for a long time, otherwise they would be caught during a later period of high wages with high-cost production techniques.

Third, it is possible that a general wage reduction would even increase unemployment. Producers would recognize that while their costs have been cut, so have their competitors', and thus might wait to see how much prices would fall before undertaking increased employment. Some might also consider the effect of the wage cut upon consumer demand, delaying employment until they could make some reasonable estimate. It is possible, too, that the first general cut in wages might be regarded only as one of several to come; then producers would delay increasing production until they were assured the wage level would fall no lower.

All these considerations lead to the conclusion that general wage cuts afford an uncertain and probably dangerous means of overcoming unemployment during generally depressed conditions. The direct cause of the economy's downswing is not usually found to be a high level of wages; other influences may be more important. And although an improved rela-

tion between costs and prices is basically needed to correct the depression, wage reduction is not the sole or necessarily a good means of effecting it. Wage cuts are of course sometimes required to correct unemployment. Wages in a particular occupation may be so high that all workers cannot be employed, whereas employment generally is good. A fall in wages for the particular type of work will help to correct this maladjustment either by driving workers into other lines and discouraging entry into it, or by increasing employment in the line through the effect of lower wages upon prices of the products produced.

If universal wage cuts are a doubtful cure for unemployment, is it likely that a general wage increase would be preferable? Unions often advance the argument that higher wages are required during a period of general unemployment to increase purchasing power. But this contention depends upon what has caused the general unemployment. Generally, unemployment is associated with a decline in the production of capital goods—machinery, plants, office buildings, etc. Production of these goods falls off because businessmen and investors consider that the prospects of profit from long-term investment are no longer attractive. The decline in profit expectations may be caused by a general decline in commodity prices, a break in stock-market prices, unsettled political conditions, a wave of bank failures, a general slackening off of consumer demand, or a combination of these and other factors. Raising wages generally could hardly be expected to offset all adverse influences, especially since a general wage increase is likely to be considered still another adverse factor by almost all employers.

A general wage increase must be judged in the last analysis by how it affects individual firms. It has already been shown that firms which are monopsonistic buyers of labor services will not reduce employment when wages are raised slightly, even though the demands for their product remain unchanged. Employers who hire labor under freely competitive conditions, however, can maintain employment in the face of a general wage increase only if the value of the marginal product of labor increases or if the combination of the productive services is not fixed. For most employers, the productive combination during any short period will be fixed—machines are designed to be run by a certain number of men if they are run at all. Such employers may, by the force of circumstances, be required to maintain employment or to close down. However, in almost any industry there are some units earning such low quasi rents that they are at the point of being closed down anyway. A general wage increase may therefore have the effect of tipping the scales in favor of a shut down and unemployment will result. This conclusion, of course, depends upon the

assumption that the general wage increase will not affect the value of the marginal product of the industry through increasing purchasing power. A general increase in purchasing power is not likely to be felt immediately, but the reduction in purchasing power caused by plant shut downs is.

It is doubtful that the general wage increase by itself will increase purchasing power in a period of considerable unemployment. The higher costs of a wage increase are specific and direct; employers are extremely conscious of this effect. An increase in purchasing power can come about only if the cost-raising effect of the wage increase is considered beneficial by the employer. Since there is no guarantee that a single employer will receive an increase in demand as a result of the general wage increase, whereas he is certain to experience an increase in costs, his demand for labor will likely decline. Moreover, if prospective purchasers of capital goods see little prospect of profit before a general wage increase, they are likely to see still less prospective profit after it goes into effect.

The foregoing conclusions on the advisability of general wage cuts or increases during a period of unemployment are in considerable measure responsible for the view, held by many economists, that the most desirable policy in depression is general wage stability, with certain adjustments in particular industries. The depression, they argue, should be attacked by direct measures to increase employment such as subsidized private investment and public work projects. By such projects purchasing power can be maintained as well as by adequate unemployment insurance.

What would be the effect of a general wage increase during a period of prosperity, when there is relatively full employment? In such a context, the effect of the general wage increase may be stimulating, since employers may be less impressed by increases in cost than by the prospective increases in demand for their products. They will, moreover, attempt to substitute other productive services for labor, and to do so many will order new machinery, thereby stimulating further the capital goods industries and increasing the demand for labor in those industries.

When labor is fully employed, increases in wages will react directly to increase prices and so wipe out the advantage to labor because of the rise in the cost of living. If prices begin to advance rapidly under the stimulus of monetary inflation, the increase in prices may actually outstrip the increase in wages. If this reasoning is valid—and many economists believe that it is—then organized labor will pursue a policy in its own and in the public interest by withholding demands for general wage increases (not accompanied by increases in productivity) when conditions of full

employment are attained. Failure to pursue such a policy in 1946 and 1947 was partly responsible for one of the most rapid advances in commodity prices ever experienced in this country.

CHAPTER 19

Interest

Interest is the price paid for the services of capital and it is measured by the rates on money loans. In an economy based upon extensive use of capital goods, this price has an influence upon the level of national output and national income.

Interest, like wages or rent, may be an explicit price paid to another for the use of capital, or it may be an implicit price determined on the basis of an explicit price which would have to be paid if capital were purchased. The term *capital* has in common usage a variety of meanings,[1] each relevant to particular problems, but in economic analysis "capital" usually refers to produced means of production. The definition of capital as "produced means of production" emphasizes the physical or "real" aspect of capital. In a modern money economy all economic goods have a monetary aspect, capital included. The purchase of capital goods requires the advance of money loans, either by the purchaser or by others; that is why the pricing of the services of capital is measured by the rates on money loans. The monetary aspect of capital later is indeed of utmost importance in the explanation of what determines interest rates, but to begin with it is well to concentrate on its real aspect.

As produced means of production, capital may take the form of buildings, machines (i.e., capital facilities), raw materials, and even stocks of finished goods used in the production or distribution of goods. Many kinds of produced means of production are also used by consumers, and yield their services directly to consumers. Thus, an owner-occupied residential building, furniture, and similar goods, although produced means of production,

[1] See Chap. 3, pp. 69-73, and Chap. 17, pp. 391-93.

are not used to turn out products for the market. These durable-use consumer goods may be called consumer capital. Many of the same goods may be producer capital under certain conditions and a consumer good under others. Coal in the stoker of a factory is producer capital; in the homeowner's coalbin it is a consumer good.

Capital is distinguished from land, which is often referred to as a "gift of nature," by the fact that capital is *produced;* from labor by the fact that it is inanimate. Two qualifications of the definition must be stated, however. First, if the capital is inexhaustible once it is produced, it can properly be classed with land for purposes of price analysis. Second, the exhaustion of the services of capital goods usually requires time. Capital goods which yield their services quickly do not tie up resources for a long period of time. Owners frequently may decide whether they wish to reinvest the money received for the services of the capital in the same kind of goods or in different goods. A manufacturer, for example, can quickly exhaust the services to him of his raw material by converting it into finished products; he cannot exhaust the services of his machinery and equipment except over a long period of time, during which his money is said to be "tied up" in capital goods. The decisions which govern long-term investments (tieing up capital in the form of specialized facilities for long periods) are therefore likely to be made more carefully and cautiously than decisions about short-term investments.

CAPITAL FORMATION IN A SIMPLE ECONOMY

The nature of real capital can be illustrated by considering a primitive type of economic life. Suppose that the natives live primarily by fishing. The state of technology is such that fishing with hook and line and with nets is known, as well as the use of boats. A youth wishing to get on in the world but possessing none of the forms of capital known to his society would have the choice of fishing in shallow water with his bare hands or making some equipment for himself. The way in which he secures his equipment, and its effect upon his output, exemplify in a vastly simplified way the process which occurs in the modern industrial economy.

A native who is earning a bare living by fishing with his hands alone—and it must be admitted that this would be an extreme case—can obtain a fishhook and line either by bartering with someone who has a hook and line or by making it himself. In either instance the cost to the native is something: if he barters for the hook and line, he must give up another

good in return; if he trades fresh fish for his hook and line, he gives up food he might otherwise have eaten himself; if he makes the hook and line himself, he gives up time which he might have spent in fishing. However, he is willing to sacrifice the fish or other result of his effort—either reducing his consumption or increasing his effort to pay the cost—in order to obtain the higher return in the future which possession of the hook and line will permit. Of course, the illustration does not demand that the native fish with his bare hands in the beginning. He might fish with hook and line, but wish to increase his catch by using a net. The requirement for obtaining a net is the same as that for obtaining the hook and line—some consumption of goods in the present must be given up in order to make or buy the net, which is necessary to increase the fisherman's catch.

The kind of production described in the case of fishing with hook and line or nets is termed *roundabout* production; and production without the use of any tools or equipment is termed *direct* production. Roundabout production consists of producing capital goods before, and for the purpose of, producing consumer goods. Since output per man is low when direct production is used, the use of roundabout methods will in most instances increase output per man. It is the superior productivity of roundabout methods which creates the demand for capital and leads people to pay for the use of capital.

Consumption must generally be reduced to enable the capital goods to be produced. Once the superior roundabout methods are put into operation, the level of consumption may be raised because of the higher productivity resulting from the "capitalistic" investment. This is the essential nature of capital formation in a simple economy. It is equally true of a modern economy when there is full employment of resources prior to a large expansion in the production of capital goods. Thus when the Soviet Union undertook to industrialize during the first Five-Year Plan after 1928, consumption was restricted because a large proportion of its productive resources was devoted to producing industrial equipment rather than consumer goods.[2] The same process occurs in wartime when consumption must be restricted in favor of "capital goods of war." Increased production of capital goods, whether for war or peace, will, when full employment is reached, require a reduction of current consumption.

[2] The reconstruction of Germany and Japan after their defeat and of other war-devastated countries will also require reduced levels of consumption.

TIME PREFERENCE—PRODUCTIVITY THEORY OF INTEREST

Thus the prerequisite for capital formation in an economy is saving, or the foregoing of present consumption so that part of current production is used for the building of capital. In the example already given, the native may "deposit" fish in advance with net and boat makers in exchange for ultimate delivery of the nets or boats. The essence of saving in real terms, then, is the foregoing of consumption out of current production. If a man receives $50 per week and deposits $5 in a savings bank, he is foregoing consumption of $5 worth of goods per week. When a corporation plows back into the business 25 cents of every dollar earned by the business and withholds this proportion of the earnings from stockholders, it is still accurate to say that this saved amount is also the result of refraining from consumption, since if all earnings were paid in dividends the consumption of stockholders might have been larger. The fact that refraining from consumption, or saving, is prerequisite to capital formation and hence greater production and consumption makes good sense of those economic virtues of industry—frugality and thrift—preached to generations of Americans and Europeans alike.

The fact that frugality and thrift are preached, however, indicates that these virtues are not inherent in the typical individual. Given the choice, most people prefer the pleasure of indulgence today to the promise of better things tomorrow; present consumption is held to be more tangible, and seemingly valuable, than future consumption. *The preference for goods in the present over the same goods at some future time is called time preference.* It may be measured by the per cent by which a given amount of money must be increased to make it as attractive at a future time as it is in the present. If, for example, a person is indifferent as to whether he had $100 today or $105 one year from today, his time preference rate is 5 per cent.

The time preference rate tends to increase as the amount of money which the individual considers setting aside is increased. Although the person will exchange $105 for $100, he may not be willing to exchange $210 for $200, and he might be reluctant indeed to exchange $1,000 for $1,050. As the proportion saved in relation to total earnings increases, the time preference rate also increases. The individual's supply schedule for saving is therefore likely to show that larger amounts would be saved only at higher rates of time preference, or *interest.* If total savings depend only upon individual

preferences, the same tendency will be true of the aggregate supply of savings.

Considering society as a whole, however, some qualifications of this tendency must be made. Some people will have a negative rate of time preference or at least very little preference for a considerable proportion of their income. For example, people who wish to provide for the education of their children or who are building up a fund in anticipation of old age or an emergency might save even if they had to pay for the privilege. Others whose incomes are very high, or who have strongly ingrained savings habits, may save almost automatically and do not value present more than future incomes.

In general, time preference of individuals is likely to be influenced by (1) size of present income, (2) the expected distribution of income over time, (3) the degree of uncertainty attached to future income, and (4) personal characteristics of the individual. A high rate of time preference is usually attributed to people with stable low incomes and with incomes which are expected to increase; a low rate of time preference is usually attributed to people with high incomes and with incomes which are expected to decrease. The effect of uncertainty of future income generally is to lower the rate of time preference if the certain income of the present can be exchanged by saving for an equally certain income in the future. Personal characteristics of the individual may modify or reverse any of these tendencies and constitute a factor which can be determined only by experience.

Actually, a large amount of saving occurs with little or no reference to the time preference of individuals. Corporations usually retain part of their earnings; these may be regarded as savings made on behalf of the owners often without their conscious consideration of the matter. Saving is involuntarily made by the millions of workers from whose pay is deducted old-age insurance premiums. As a result of the economy's price system, saving is forced by the vagaries of price movements. As the level of prices rises, people with constant incomes (in terms of money) are forced to curtail their consumption. And finally, as a result of human tastes, many people with high incomes save more or less automatically because their incomes are far in excess of the amounts required to satisfy their wants. In the moderately prosperous year 1937, 31 per cent of total savings were made by business enterprises, and 9 per cent by governments. A little over 59 per cent of the total was saved by individuals.[3] Over 39 per cent of the individual savings in 1935-36 was made by those families and single

[3] Oscar L. Altman, *Saving, Investment, and National Income*, Temporary National Economic Committee, Monograph No. 37, 1941, p. 110.

individuals who received an income in excess of $20,000 per year.[4] Little of the saving by business and by government could be said to be influenced by time preference, and the saving of individuals in the high-income groups was probably but little influenced by time preference.

If the capital requirements of the economy are small, it is possible that the savings of business and of certain individuals whose time preference is negative would be large enough without the payment of interest. In ordinarily prosperous times, however, it is likely that the supply of voluntary savings at a zero rate of interest would not suffice to meet the demand at that rate. Increased savings could thereupon be induced by paying a positive rate of interest on savings large enough to overcome the positive preference of persons for present consumption. It will be seen, however, that the amount of saving is likely to be influenced more by the level of national income than by the level of interest rates.

Saving and Investment

It is clear that capital goods can be secured only if some people are willing to reduce their consumption to "finance" the making of them. The "finance" in a simple economy, such as the hunting and fishing one described, consists of supplying consumer goods to those who make capital goods. Even in a money economy the real aspects of finance are the same. When, for example, money is advanced to finance the construction of a building, in real terms the money advance represents claims on material needed for the building and consumption goods purchased by workers and others who erect it.

It follows, therefore, that there is a direct relationship between saving and investment, which is another name for capital formation. The individual's disposable income is used for consumption and for saving; his output is made up of consumption goods and investment goods. Since for an individual in a simple economy income and output are the same thing, saving and investment are equal. Although it is more difficult to understand, this same relationship between saving and investment is true for any period of time for any kind of economic society, simple or complex, taken as a whole. This was shown in Chapter 4; gross saving equals total investment expenditure, and net saving equals net investment.[5] The actual

[4] *Consumer Expenditures in the United States, 1935-36,* National Resources Committee, 1939, p. 48.

[5] See pp. 101-05, also 119-23.

process by which equality of savings and investment is maintained is the most controversial aspect of the theory of interest.

In a simple fishing economy the process by which saving and investment would be equated is simple. For the individual fisherman, increased saving is likely to be accompanied by a rising time-preference rate; on the other hand, the opportunities for investment of a poor fisherman may be great. By using a hook and line he may be able to double his daily catch, and by using a net to triple or quadruple it, and by using a boat to increase it still further. The fisherman, obviously, will seek to adjust his saving and investment to an optimum position, which will be reached when the return from additional investment equals the price (time-preference rate in this case) of the additional saving necessary to finance it.

Thus for a simple economic society in which investment is directly dependent on the supply of savings, investment and savings are equated by means of a rate of interest, which is a time-preference rate. If money is used as a *measure of value,* for purposes of illustration, various amounts of saving (measured in money) would be forthcoming at various rates of interest and the amount of investment (measured in money) would be governed by various rates of interest. Assume that all loans are of only one type for the present. Considering only the rate of interest for such loans, it is obvious that during any period of time, saving in excess of a certain amount will not be made at all unless some interest rate is received. The higher the rate, the greater is the amount of savings likely to be. Because of the influence of the law of diminishing returns, increased investment would be made only at lower rates of interest. The demand for savings and the supply of savings will be equated at some rate of interest. Interest, the price for the services of capital, in this case brings about the inevitable equality of the amount saved and the amount invested and determines what this amount will be.

The forces determining the rate of interest in a modern economy have been in recent years the subject of considerable argument among economists. The time-preference-productivity theory holds that savings and investment are equated by the rate or rates of interest. The rate of interest is regarded as a price which equates the supply of savings—governed by time preference—with the demand for saving—governed by the marginal productivity of capital. According to this theory, if people attempt to save at current interest rates more than will be invested at those rates, interest will fall. Although it is plausible under conditions which prevail in a simple economy, this theory neglects the fact that savings and investment cannot be considered to be independent. In reality, savings and investment

do affect each other. For example, when investment is increased, national income will increase and savings also will increase. It will also be found that the demand for capital can, to a considerable extent, create its own supply of saving, independently of any change in interest rates.

LIQUIDITY PREFERENCE THEORY OF INTEREST

In view of this consideration, an alternative theory of interest—the *liquidity preference theory—maintains that the forces determining interest rates are not saving and investment but the supply and demand for money loans.* This theory is termed the *liquidity preference theory* because the supply of money loans depends upon the supply of money, or liquidity. Since the supply of and demand for money loans are held to determine interest rates, savings and investment are relegated to a position in which they adjust themselves to the interest rates established. The liquidity preference theory introduces the effect of certain forces in the economy which are possibly overlooked by the time-preference-productivity theory.

The liquidity preference theory rejects the idea that investment is *necessarily* dependent upon the willingness of people to save, as in a simple economy. The amount of investment is dependent upon the willingness to save only under conditions of relatively full employment; when unemployed resources are at hand investment may be increased without any reduction in the previous level of consumption. Indeed it may increase to the accompaniment of a rise in consumption.

For example, suppose that every city, town, and village in the United States, in response to a request from the President, undertakes during a year of large-scale unemployment to invest $100 per head of population in public work projects, with the funds provided by a special issue of money by the Federal government. During the course of the year, several billion dollars would be invested in community facilities, roads, and other public improvements, and employment would increase (assuming that the special issue of money had no adverse effects on business confidence). The rise in national output, national income, employment, consumption, and savings occasioned by such a policy has been described in Chapter 6.[6] As income increases, consumption usually increases also, so that on balance the new investment (i.e., public works) may be found to be accompanied by increased rather than by reduced consumption. How, then, can the additional investment be balanced by additional saving, since saving must always equal investment?

[6] Pp. 184-89, also 142-44.

Equality of saving and investment for the period is achieved either through increased saving or reduced investment of other kinds or a combination of the two. Storekeepers who have increased their sales may not have increased their orders in proportion; at the end of the year their inventories will be reduced; they will have *disinvested.* Their incomes, on the other hand, will be increased; and even if they spend the same amount as before, they will have increased their savings. Also, the newly employed may make some savings, and those who supported them when they were unemployed are likely to save more. Since this trend is general the net result will be that the new savings are equal to the increased total investment. Significantly, this process of equating saving and investment can be described without any reference to interest rates, even though they may be affected themselves by it.

The new elements in the problem which were not present in the simple economy are (1) unemployed resources and (2) money as something more than a common denominator or measure of value. The existence of unemployed resources makes it possible to increase investment without at the same time reducing anyone's consumption. Increasing the quantity of money may make it possible to provide the funds for new investment without reducing anyone's spending. The example given is oversimplified, especially since it is assumed that the new investment can be purchased with printing-press money, but it brings to the fore the dominance of money and unemployment as factors affecting the relation of interest to savings and investment.

In the modern money economy there are no markets where savings and investment meet. There are only markets for money loans. When a person wishes to borrow, almost always he wishes to borrow money; to do so he takes a loan through a bank or other lending institution. If banks lent only the savings left with them, and if the loans were all invested, then a complete examination of the supply and demand for loans would in effect constitute a study of the market where savings and investment meet.

Actually, however, the supply of money loans is only remotely related to the supply of savings. The commercial banking system is responsible for the major part of the supply of money in the country; about 90 per cent of the country's payments are made by drawing upon check deposits in the banks. Commercial banks usually lend money by setting up new deposits in favor of the borrowers, on the basis of the borrowers' promises to pay backed up by such security as is mutually agreed upon. But the banks' ability to lend is not limited to the deposits of cash left with them by the public, for the banking system as a whole has the power to create

deposits in exchange for promises to pay and other securities. The supply of loans, therefore, depends very largely upon the *willingness of the banks* to create deposits, or, what amounts to the same thing, upon their willingness to hold securities and other assets.

The willingness of the banks to hold securities and other assets depends not primarily upon the rate of interest but largely upon business conditions and the state of bank reserves. Every deposit on a bank's books represents a potential demand for cash. Since cash earns no interest, banks usually attempt to hold as large a proportion of their assets as possible in the form of securities and other earning assets. They are, of course, required by law to maintain certain cash reserves. If both the banks and the public feel that business is good and getting better, the banks may be willing to increase their loans by creating new deposits. If the banks—or the banks and the public—feel that business conditions are uncertain, or that they are almost certainly going to be worse, then the willingness of the banks to lend diminishes. They will not increase the volume of loans; rather, they will try to accumulate cash, in order to meet such emergencies as a general withdrawal of cash for hoarding by depositors, or adverse clearing balances with other banks.

From these considerations, therefore, it can be stated that *the supply of loans is dependent upon the liquidity preferences of the banks and the public.* Liquidity preference is the preference for cash, the most liquid of all assets, over other assets such as promissory notes, stocks, bonds, or real-estate mortgages.

The liquidity preference of a bank can be satisfied only with the kind of money which will meet the bank's obligations to other banks and to the public. A bank's obligations to other banks arise because it has itself borrowed from other banks or has accepted deposits of other banks placed with it, or because it must meet the payments of checks drawn on it which have been deposited with other banks. These interbank debts can be settled only with money generally acceptable to banks. This money is, ultimately, either claims [7] upon the Federal reserve banks or currency.

The liquidity preference of the public can usually be satisfied either with demand deposits or currency. However, in times of great uncertainty the confidence of the public in the banks diminishes and many people demand currency.

Liquidity preference of the public arises from three motives for holding money. (1) Money (usually demand deposits) is held because receipt and disbursement of income of both individuals and businesses are not simul-

[7] See Chap. 10.

taneous. The larger the volume of the national or individual income, the larger the volume of money likely to be tied up in bridging the gap between receipts and expenditures. (2) Money is held to provide for unforeseen contingencies, such as being called upon to pay debts or to meet a sudden emergency like an accident. (3) Money is also held for speculative purposes. When, for example, a person expects the prices of some of his assets to fall he will try to convert them into cash and will hold the cash until he is convinced that the market for those assets will fall no further or will possibly rise. One thousand dollars in cash is always $1,000, but $1,000 worth of United States Steel common stock today may be worth $800 next month.

When both the banks and the public show a marked increase in liquidity preference, banks try to reduce their loans and investments in order to increase their reserves and claims on other banks, and the public either holds bank deposits without using them or actually converts its bank deposits into currency. The net result of a dominant liquidity preference is a dumping of assets on the market.

The supply of loans is inversely related to liquidity preference: when the liquidity preference of the public and the banks rises, the supply of money available for loans tends to fall. However, liquidity can be purchased only at a price—the price of interest sacrificed by holding money rather than those assets which bear interest. Banks and other lenders are more willing to convert money into other assets (i.e., make loans) at higher than at lower rates of interest. Thus, the volume of loans is dependent at any one time upon the willingness of lenders to give up liquidity.

Just as the supply of loans is not necessarily dependent on saving, so the demand for loans is not necessarily dependent upon investment. Borrowing may be either for consumption or for investment. The demands for investment loans, which predominate in normal times, depend upon the value of the marginal product of the capital secured by means of the loan. If a farmer expects that by borrowing $100 for fertilizer he can add $125 to the value of his next crop, other things remaining the same, then a loan of $100 will yield a return of 25 per cent. If loans can be secured for 6 per cent, it would clearly pay to borrow $100 for fertilizer. If a second $100 applied to fertilizer would yield 15 per cent, it would still pay to borrow a second $100 at 6 per cent. It will pay to borrow funds and use capital up to the point where the last unit yields as great a price as its cost, *or where the prospective marginal rate of return equals the rate of interest.* If the farmer finances his own operation, the principle is the same, for the farmer may be considered to be lending to himself. (On the other hand,

the demand for loans for consumption at any time is also influenced by the rate of interest but consumers are concerned with their own time-preference rates rather than with the marginal productivity of capital. In a nation of spendthrifts, the rate of interest would be influenced more by the demand for consumption loans, both private and public, than by the marginal productivity of capital.)

There are many types of investment, each yielding income in a different manner. The simplest type, such as the one just illustrated, is that which is expected to return its cost and its income in one year's time. Still another type is that which is expected to yield its return perpetually, e.g., an investment made to drain land for farming. Others, like investments in reforestation or fruit-tree planting yield nothing for a time and then yield an income over a period of years. Still others, like investments in buildings or machinery, may yield income at once and steadily over a period of years. Although these cases are more complicated than simple one-year loans, the principle applicable is still the same: equilibrium for the investor is achieved when the marginal rate of return is equal to the rate of interest. (The equilibrium rate of return in the case of a stream of income is that rate which yields a present value of the prospective receipts equal to the value of the investment required to secure it.)

Thus far it has been shown that interest rates are both influenced by and themselves influence saving, investment, the demand for money loans, and their supply; and that the supply of loans depends upon the liquidity preference of the public and the banks. It is only for convenience, therefore, that the supply of and demand for money loans are considered to determine interest rates, since these forces are not independent but are the result of many other forces operating in the economy. The equilibrium structure of interest rates in the modern economy equates (1) the demand for and supply of money loans of all types, and (2) the preference for holding other assets with the preference for holding cash. These relationships in which the rates of interest have a predominant role lead to some conclusions about a modern economy which are strikingly different from those applicable to a simple economy.

Suppose, for example, that in a simple economy investment falls off. The rate of interest will fall, people will save less and consume more, and a new lower level of interest rates will result. On the other hand, if investment falls off in the modern economy, the reaction is not so direct or simple. As investment falls off, the demand for loans will probably fall. If the reduction in the demand for loans is looked upon by the banks and private investors as a sign of approaching bad business conditions, both investors

and banks will probably undertake to become more liquid, reducing their loans and investments to get money. If this movement toward liquidity is general enough the level of interest rates may rise instead of fall, because an increase in liquidity preference results in a decrease in the supply of loans. People may attempt to save the same amount as before, especially if the level of interest rates does not change or even if it rises. But the economy as a whole cannot increase its savings, because investment has fallen. The effect of maintaining savings must be to reduce incomes of others in the economy, since the decline in investment means a reduction in income of those whose income is dependent on the manufacture of capital goods. Finally, then, if savers attempt to maintain or to increase their saving, the net effect upon the economy as a whole must be a combination of (1) reduced saving and more dis-saving (drawing down of savings) on the part of those whose incomes are reduced (such as workers out of jobs and shopkeepers with reduced business), and (2) involuntary investment on the part of those who bought goods expecting to sell them but now find them left in inventory. A decline in investment in the modern economy, therefore, does not necessarily bring about a reduction in the rates of interest, just as an increase in investment does not necessarily bring about an increase in interest rates.

A COMBINATION OF THE THEORIES OF INTEREST

All persons may use the money in their possession for any or all of the following purposes: (1) to buy commodities, (2) to buy securities (to make loans), or (3) to satisfy their desire to hold money as such. Whether or not they buy commodities is partly influenced by their time preferences. Whether or not they buy securities is greatly influenced by their willingness to give up liquidity—demand deposits or hoards of currency. At any given time, there must be a balance among these uses for money.

Prices serve to maintain this balance. At any level of the national income and of interest rates, the demand for and supply of goods and services will be equated by their prices. The amount of money people will hold to satisfy transactions and precautionary motives is likewise determined when the prices of goods and services are established, since this amount is dependent upon income and all prices but interest. For each possible level of national income there is (1) a schedule of amounts of money people are willing to lend at various rates, and (2) a schedule of amounts which people are willing to borrow at various rates (on the basis of productivity calculations for production loans and time-preference schedules for consumption

loans). Interest rates therefore are not determined by the amount of money people use for dealing in goods and services, but by the supply of and demand for money loans. The amount of money people do not wish to devote to transactions and precautionary motives becomes the source of supply for money loans. Supply and the demand for such loans thus determine interest rates.

The Structure of Interest Rates

Up to this point in the discussion it has been implied that all interest rates move together; rates on different types of loans have not been distinguished. If all loans were riskless and cost nothing to make, there would probably be only one interest rate. In reality, there are many rates in the loan markets, and these depend, among other things, upon (1) the security pledged, (2) the length of loan, (3) the degree of liquidity of the loan, and (4) the type of borrower. The separate effects of these various considerations upon interest rates are almost self-evident; however, they are all at work together so that the effect of any specific one cannot be determined easily.

Two types of loan identical in every other respect may carry different rates of interest if the borrower for one has better credit than the other. Thus, as between a 20-year railroad bond of the highest type and a 20-year United States government bond, one would expect the latter to carry a lower rate of interest. The risk of default on a bond issue of the United States government is nil.

The length of the loan contract also influences the rate of interest. A long-term loan carries considerably more risk of depreciation in value should the level of interest rates change than is true in the case of a short-term loan. Suppose one has paid $1,000 for a 20-year bond paying 2 per cent interest, or $20 per year. If the rate of interest paid on similar bonds issued some years later should be 3 per cent, then for $1,000 it would be possible to purchase an income of $30 a year. This would cause the market value of the 2 per cent bond to fall until it yields 3 per cent; i.e., its price would fall to about $667. Hence, the holders of long-term securities are always subject to the risk of unfavorable changes in the market price of their securities if the rate of interest rises above the contract rate. In contrast, a "call loan," one which may have to be repaid on a few hours' notice, carries virtually no risk of depreciation because of changes in interest rates, because it can be called and reloaned at a new rate almost immediately. The short-term lender may restudy the financial position of the

borrower, the adequacy of the collateral, the degree of liquidity of the loan, the state of the loan market, and other pertinent facts at frequent intervals. Therefore to induce lenders to buy longer term securities, borrowers must offer higher rates of interest.

Furthermore the liquidity of loans influences interest rates and sometimes results in extreme differentials. Other things being the same, a more saleable security carries a lower rate than a less saleable one. Thus, given two loans with the same length of term and same degree of risk, the one having the readier market will tend to carry the lower rate.

Finally, interest rates must be considered to include the costs of lending, since virtually all interest rates include costs other than "pure" interest. Banks maintain buildings, pay salaries to officers, investigate the credit of potential borrowers, and, in general, they must meet the expenses of an operating business enterprise. The rates charged must of course include these expenses if lending is to be carried on. Likewise, in the buying and selling of securities some costs of the transactions are borne by brokers or other intermediaries; these affect the interest rates paid or received by borrowers or lenders, and often are taken into account in the bid and asked prices of the securities. In the small-loan business, these costs of lending—maintaining offices, investigating clients, keeping accounts, collecting from clients whose payments are overdue, and taking possession of collateral—contribute to the high rates of interest charged. The distressed financial position of most of the customers of small-loan businesses, on the other hand, has encouraged many lenders to exploit their need by charging excessive rates, a consequence of which in most states has been the enactment of legislation limiting personal-loan rates.

The Government and Interest Rates

The supply of loans, it has been shown, is determined by the willingness of lenders to offer money for securities. If nonbank lenders are amply supplied with money and if commercial banks have large excess reserves, their willingness to lend should be strong, since money yields only liquidity, whereas securities yield some income in the form of interest. In recent years the government has directly or indirectly dominated the loan markets. It has achieved this by pursuing a policy which has kept the banks adequately supplied with reserves.

The mechanism by which the Treasury and the Federal Reserve System can keep the banking system and business community in a highly liquid state has been explained in the chapters on money and banking. The

government has seen to it that the banks have been in a position to meet all demands for loans. The Federal Reserve System, in addition, supports the market for government securities by standing ready to buy at any time when prices fall too low, thus maintaining low interest rates. So powerful and pervasive has been the government's influence in recent years that there is not much doubt that it has dominated the level of interest rates.

Interest and Capitalization

When the supply of money available for loans increases, other things remaining the same, the rate of interest will tend to decline. Since the supply of money available for loans is simply the demand on the part of lenders for securities, an increase in this supply is likely to raise the price of securities.

The value of any income-bearing right is dependent upon the going interest rates. If a machine is estimated to yield $50 per year (over and above the amount necessary for maintenance and depreciation) its value can be determined by equating this yield with the interest rate on designated securities. If a security paying $50 per year for the same period as the estimated longevity of the machine is selling for $1,000, then the machine may be valued at $1,000 on the basis of the income it is expected to yield. It is not practicable, of course, to ascertain the market value of a security comparable to every property right to receive income; and for a great many rights no comparable security may be found on the market. Nevertheless, it is possible to treat every income-bearing asset as if a security comparable to it were priced by the market. Suppose, for example, that a person held a two-year contract for a small downtown parking lot estimated to yield a net profit of $1,000 per year. If he wished to convert this two-year contract into money at once, he might sell the contract to someone else. How much would it be worth? The value of the contract would depend, assuming the return of $1,000 per year to be virtually a certainty for two years, upon the return the buyer requires on his investment. If the buyer were not content with less than 10 per cent, then the value of two $1,000 payments—one at the end of each of the two succeeding years—could be calculated. A sum of $909.09 invested for one year at 10 per cent will yield $90.91, and a sum of $826.446 invested at 10 per cent compound interest for two years will yield $173.554. The sums invested plus the interest yield amount to $2,000. Therefore, if the buyer wishes to earn 10 per cent on the right to receive two $1,000 payments, one in one

year and one in two years, he will pay $909.09 plus $826.45, or $1,735.54 for the privilege.[8]

In actual practice, the purchase of an income-bearing asset is not simply a matter of determining what rate of interest should be applied to the sums due in the future. Purchasers of business property, securities, or any durable agents of production must take into account not only the appropriate interest rate, but also must consider the risk that the expected future payments may materialize only in part or not at all. Interest return encourages investment but is often outweighed by fear of lowered rates or business failure.

The process by means of which present values of payments due in the future are linked with interest rates provides the mechanism for connecting the time preference of individuals with the prices of income-yielding assets. The higher a person's time preference, the higher will be the implicit or explicit rate with which he discounts (finds the present value of) payments due in the future, and the lower the price he will offer for rights to receive payments in the future. In other words, higher time preference seeks higher future returns and lower present prices for rights to income.

The Function of Interest

To the borrower who must pay interest on his loan, interest is a burden which he would prefer to eliminate. Yet interest, like any other price, performs a function of allocating economic resources. Interest allocates expenditures between consumption goods for current consumption and capital goods which serve to increase consumption in the future. When resources are used to make capital goods, the capital goods are expected to yield a net return equal to or greater than the going rate of interest. Thus, while current consumption is restricted by the process of capital formation, a greater supply of consumer goods is made available in the future than would otherwise have been possible.

Moreover, interest also restricts the production of capital goods to those types which will yield a relatively great supply of goods in a short period. Between two machines, each costing (having a capital value of) $10,000, and each capable of producing the same value of products, one may produce this value within 1 year while the other may produce it only over a

[8] The process of finding a present value for future payments uses the compound-interest formula $P = S\frac{1}{(1+r)}n$. P is the present value sought, S is the future payment, r is the interest rate, and n is the number of years which must elapse before the sum is payable.

period of 10 years. The cost of interest thus dictates the selection of the former machine.

If interest rates are high, only those capital goods will be built which, for a given state of technology, result in the greatest increase in productivity in the shortest possible time. The volume of capital formation—and the proportion of resources devoted to the production of capital goods—will therefore be limited. At high interest rates only capital goods of high productivity will be produced, whereas low interest rates permit the building of capital goods which increase productivity only slightly. Since there are always more opportunities to use such goods, the total amount of capital formation or investment at low rates of interest is likely to be greater than at high rates.

The different interest rates also play a part in allocating loans among different types of borrowers. If people prefer to anticipate income by buying automobiles, homes, and other goods on installments, the rates on loans for these purposes will tend to rise and lenders will be attracted to this type of business and perhaps away from financing the production of producer goods. Or, if lenders indicate a strong preference for safe, long-term securities, the demand for securities of this type will result in lower interest yields (higher prices) for this type of security.

SUMMARY

Capital, which consists of produced means of production, is demanded because it increases the productivity of the combination of factors in which it is used. Capital formation, however, requires under conditions of full employment some restriction of consumption below what it otherwise would be. People are willing to restrict consumption voluntarily, or to save beyond a certain point, only if they are promised a premium in the future for goods sacrificed in the present. This is because most people have a positive time preference, which is a preference for present goods over the same goods in the future.

Both time preference and the productivity of capital are measured by interest rates, which are expressed as a per cent per period of time on the capital sum (almost always measured in money).

In a modern economy, interest rates are the prices of money loans established in the markets for such loans. Interest rates are, therefore, determined primarily by the supply of and demand for money loans. The supply of such loans depends not upon saving, but upon the willingness of lenders to give up liquidity by making loans or buying securities. The de-

mand for loans is determined by the marginal productivity of capital at the level of national income prevailing or expected, and by the time preferences of individuals seeking consumption loans. The level of saving depends upon the level of national income much more than upon interest rates, although individual savers are undoubtedly influenced by interest rates.

Interest rates on different types of loans are influenced by such factors as liquidity, security, length of the loan, and cost of lending. Because these aspects of different types of loans have unequal effects, the structure of interest rates in an equilibrium state will show different rates for different types of loans. The yields on old loans will be adjusted to those on new loans of the same type by changes in the capital value of outstanding securities. The capital value of securities outstanding will fall when interest rates for new securities of the same type rise, and the capital value of old securities will rise when the interest rate on new securities of the same type falls.

The primary function of interest rates is to ensure that the future output of goods resulting from the use of capital is sufficiently increased to compensate for the restriction of consumption which may be involved in the building of capital goods. Varying rates for different types of loans serve to channel loans into those lines most preferred by consumers, as evidenced by their demands for consumer goods.

CHAPTER 20

Rent

The prices of the productive services of resources which are, for any problem under consideration, fixed in supply are called rents. Rent by this definition departs from the common usage, as in "renting" a house or apartment, a dinner jacket, or an umbrella at the beach, in which rent is taken to be a periodic payment for the hiring of some object. However, this ordinary meaning of rent is inadequate for classifying the prices of productive services *in terms of the supply characteristics of the sources which yield them.*

One may rent an apartment for which the lease contains the guarantee that maid service as well as electricity, water, and gas will be supplied by the management. The rental then includes costs for public-utility charges, and for other goods and services. Out of it must come part of the salary of the manager, part of the taxes, some of the maintenance cost of the building, and possibly some of the interest and amortization of the investment, to say nothing of profit for the entrepreneur-owners. It is obvious that the productive services compensated by payment of the rental are not simply the services of resources in fixed supply. Furthermore, ordinary usage makes a sharp distinction between owned and rented property. The man who lives in a house which he owns seldom regards himself as paying rent on it, yet for many purposes it is desirable to impute a rent to property used by the owner. (This is now done in making estimates of national income.) For purposes of economic analysis, therefore, the ordinary meaning of rent is insufficiently defined: "rent" is used here as the price of resources which are in fixed supply.

In the development of economic studies, the modern theory of rent typically received its greatest impetus during a period of economic upheaval. The Napoleonic blockade restricted the shipment of grain to Britain and grain prices and therefore the price of bread rose to high levels, causing widespread hardship. The price of wheat in England rose rapidly after 1794, and by 1801 had reached three times the 1794 price. For a number of years thereafter, the price of wheat remained abnormally high in relation to its prewar level. Many English reformers and agitators attributed the increase to the avarice of the landowners who raised their rents.

David Ricardo, a businessman turned economist, published in 1817 his *Principles of Political Economy and Taxation*, setting forth among other things the so-called Ricardian theory of rent. Ricardo maintained that rent was not the cause of high wheat prices, but, on the contrary, rent was the result of these prices.[1] This conclusion was partly responsible for the repeal in 1846 of the Corn Laws, which had been Britain's instrument of agricultural protection. The British market was then opened to cheap grain from abroad. An examination of the causes and consequences of rent, the topic which so excited early nineteenth-century England, can proceed from this point.

Land is traditionally regarded as the source of productive services fixed in supply. Although this was never strictly true, land nevertheless repre-

[1] Although Ricardo is generally credited with developing the rent doctrine, he was in fact preceded by a Scotch farmer named Anderson in 1777, and by T. R. Malthus and David West in 1815.

sents a source of productive services which in the long run comes closest to being fixed in supply. The land area of the world is virtually fixed in amount, and the additions to it have been relatively insignificant. However, even in the long run not all sources of productive services which are fixed in supply need be land, and all land is not necessarily fixed in supply. Improvements to land are made which are almost as permanent as land itself, such as drainage systems, sea walls, etc. These should properly be considered to be a part of the land itself, once they are made. On the other hand, land is subjected to wastage through overcropping and erosion: thus, it is incorrect to regard land in all its aspects as a fixed source.

In the analysis of long-run economic problems, the term *rent* may be considered to be the price of the services of land, after deducting expenses required to maintain it in what is held to be its original condition. If the price is actually paid to the owner by the user, he receives *explicit rent*, and if the owner uses the land himself, he receives *implicit rent*. The amount of implicit rent is the price which would be paid to the owner for the services of the land alone if it were actually leased to another user.

Although the outstanding characteristic of land is its fixed supply, which makes its services available whether any price is paid for them or not, rents are nevertheless dependent upon the forces which are central to the marginal productivity theory. If land is scarce in relation to the demand for it, it must be used economically. To economize on the use of land, one must (1) choose that use for any given piece of land which yields the highest return, and (2) in that use, combine the land with other factors up to a point where the value of the marginal product of the land is equal to the price of its services, or rent.

To illustrate the economical use of land, suppose that the number of units of land cultivated is varied in proportion to a given combination of labor and capital (men, machinery, and fertilizer). The cost of labor and capital is assumed not to vary with the type of product (oats, wheat, and corn) produced. In accordance with the law of diminishing returns, as more land is used with a given amount of labor and capital, the value of the marginal product falls; for the greater the amount of land per unit of labor and capital, the less effective will be the cultivation.

Under the returns assumed in Table 45, the growing of wheat yields the highest returns; however, it is not clear which *amount* of land in combination with the labor and capital would be most advantageous. If good farm land were so abundant in a given location that it could be had only for the taking, then the farmer would maximize the return of his labor and capital by using large quantities of land in proportion to them. In the

present example, he would use up to six units of land and so he would equate the price of land, which is nothing, with its marginal product. However, if the land were so scarce that some people would be willing to pay for the use of it, the equation would change. If the prospective farmer offers anything over $50 for the services of one unit of land, the owner will rent it to him, since by giving up one unit of land the owner's total product is reduced only by $50. If all land were of the same quality and all the people who wanted to farm could obtain at $50 per unit all the land which they want at that price, then $50 per unit would be the price for the services of such land. Supply would be equal to demand at that price.

Table 45 **VARIATIONS IN THE PRODUCT OF LAND WITH DIFFERENT COMBINATIONS OF FACTORS**

Number of units of land used with fixed amount of labor and capital	Return in oats		Return in wheat		Return in corn	
	Value of total product	Value of marginal product	Value of total product	Value of marginal product	Value of total product	Value of marginal product
1	$125	—	$250	—	$187.50	—
2	225	$100	450	$200	337.50	$150.00
3	300	75	600	150	450.00	112.50
4	350	50	700	100	525.00	75.00
5	375	25	750	50	562.50	37.50
6	375	0	750	0	562.50	0

It makes little difference whether the demand for land is met by renting or purchase. If land were not scarce, no amount would be paid for it. If it is scarce and people are willing to pay something for its services, they may do so by paying an explicit rent for the use of land, or they may purchase the land outright, paying a price for title to the land which reflects the explicit rent that would have to be paid otherwise.[2]

The foregoing discussion of rent has a number of applications of considerable importance. It will be observed that rent comes into existence because land is scarce relative to the demand for it; the existence of rent makes titles to land valuable. The course of American history dramatically verifies this fact. The United States began its national history under conditions in which land was the principal economic asset. As the eastern seaboard increased in population, the demand for land rose and made it more valuable. People who wanted land for farming—and until comparatively recent times the country was predominately agricul-

[2] See Chap. 19, pp. 443-44.

tural—had the alternative of renting or buying land in the east at high prices, or moving westward to buy up cheaper land or homestead on virtually free land. Moreover, people with money available bought land and sold it later as the country filled up at greatly increased prices. Thus an almost irresistible economic force drove the nation westward, a force which such minor obstacles as Indian troubles and the loneliness and hardships of life in an unsettled country did little to check.

Because of the great amount of cheap land, agricultural practices grew up which amazed European travelers in this early period and which even today stamp American farming as unique in some respects. The most economical combination of productive services was, because of much land and small cost, one in which the services of land were large in proportion to the services of the other factors. Travelers constantly commented on the "wasteful" practices of American agriculture, particularly in newly settled areas: stumps were not removed from newly cleared land, plowing was haphazard, natural manure was allowed to go to waste, and little attention was paid to preserving or improving the soil. In consequence, crop yields per acre were often small in comparison to the yields from more intensive European cultivation—but the yields per man were higher.

The "wasteful" practices of early American agriculture can be explained in terms of cheap land and expensive labor: relatively more land and less labor were used in the combination because it paid to push the use of land to a lower point on the marginal productivity schedule, the price of land being lower here than abroad. Even today, western Europe produces a higher physical yield per acre in basic grain crops because it employs more labor per unit of land, the services of land there being higher in price relative to the prices of labor and capital services. Rents are much higher relative to wages in western Europe than they are in the United States.

DIFFERENTIAL RENTS

It has thus far been assumed that all land is uniform in productivity. But all land, even agricultural land, is of course not the same. Different pieces of land vary in fertility, in topography, in location relative to markets, and in many other ways. Because of these differences, the productivity and therefore the rents of various pieces of land are different.

For purposes of explanation, it may be assumed that there are four grades of land in a given area which differ only in fertility and that they are all most advantageously used for growing wheat. Suppose that a variable

number of units, each consisting of a given amount of labor and capital, are used with a fixed amount of each of the four grades of land. The results are shown in Table 46.

Table 46 DIFFERENCES IN THE YIELD FROM LAND

Labor-capital units applied to a fixed amount of land	Value of Marginal Product			
	Grade I land	Grade II land	Grade III land	Grade IV land
1	$50	$40	**$30**	$25
2	45	35	25	20
3	40	**30**	20	15
4	35	25	15	10
5	**30**	20	10	5
6	25	15	5	0
7	20	10	0	− 5
8	15	5	− 5	−10
9	10	0	−10	−15
10	5	−5	−15	−20

In this example the price of wheat is assumed to be given, and the price per unit of labor and capital is set at $30. Up to five units will be used on grade I land because the value of the marginal product exceeds $30—the price of the unit of labor and capital—up to the point where five units are used. Three units of labor and capital will be used on grade II land, and by the same calculation only one on grade III land. It will not pay to cultivate the poorest grade of land at all because less than the outlay of $30 is returned for even the first unit of labor and capital. The total cost of labor and capital services is $150 on grade I land, the total return being $200. Considering only the use of labor and capital on the land, the difference of $50 between cost and return can be attributed to the land as rent. Likewise, the rent on grade II land will be $15. Grade III land, which is just able to return the cost of labor and capital used on it (one unit at $30), yields no rent although it is cultivated. Grade IV land will also yield no rent, because of course it cannot be cultivated without loss. In this way, then, differences in the productivity of land give rise to differences in rent. It has become customary to identify that grade of land which it barely pays to cultivate (grade III land in this example) as the *extensive margin* of cultivation, and the last unit of productive services which it pays to apply to land of better quality the *intensive margin* of cultivation (the fifth unit of labor and capital used on grade I land, and the third on grade II land in this example).

If the price of wheat should rise, the value of the marginal product of capital and labor employed on all grades of land would increase. Assuming that the price of capital and labor remains the same, more units of capital and labor would be used per acre of land; and the rents would increase. As rents increase, the purchase price which must be paid for the land itself (the capitalized value of the rents) also increases. Moreover, if the value of the marginal product of grade IV land is raised to $30 or more, it will now become desirable to cultivate it. When the prices of agricultural commodities remain high for any considerable period of time, such land will be brought into cultivation. This process was observed during both world wars. However, if landowners attempt to charge tenant-farmers more than the differential return from the land they cultivate over the return on the extensive margin, tenants have the option of moving to no-rent land. The existence of such land is, of course, obscured by the fact that something must almost always be paid by tenants for farm lands of even the poorest quality, but such payments often consist of nothing more than a price paid for the hire of buildings, tools, and equipment furnished by the landowner.

Perhaps the most striking illustration of differential rents is found in urban rents, which arise primarily from differences not in the fertility of the land but from its location. The concentration of people in downtown areas, at least during the daytime hours, makes certain areas particularly attractive for business purposes. The number of persons passing various store sites has a direct bearing on rent differentials in a retail business area.

Sites in certain districts in cities have high values because of the large volume of business which is transacted, proximity to transportation facilities, and because of the limited number of sites with such advantages. The most valuable feature of such business properties is *space* in a particular location, and since space is limited, every effort must be made to substitute relatively more of other productive services for the services of land. Thus the skyscraper represents an extreme substitution of labor and capital resources for land. For buildings located on a limited and valuable amount of land, many additional floors can be built before the value of the marginal product of capital (building facilities) is reduced to the cost of such capital. Diminishing returns will apply to the application of capital to land in this way, however. As buildings extend higher a given amount of capital will provide progressively less space because of pyramid-form construction; furthermore, space on all floors is reduced and costs for each floor increased because heavier foundation, walls, and frames and the installation of elevators is required after certain levels are reached.

Downtown districts in cities are generally the most valuable, attracting stores, office buildings, theaters, and hotels, businesses which have the largest earning capacity from the space available. In many cities, the earning capacity of some apartment houses is apparently next after office and store buildings, and indeed such houses are in large part valuable because of their nearness to downtown stores, offices, and theaters. Farther out from the downtown districts, located in conjunction with highway, railroad, and water transportation portals, are found warehouses, light manufacturing industry, and then heavy manufacturing. Land in these districts of the city yields considerably more rent for these purposes than for residential use or for farming. Beyond the manufacturing district, and often on the side of the city opposite from the principal manufacturing area, are found residential areas which bring higher returns in this use than in farming. These residential areas merge into the suburbs, where truck gardens and nurseries specializing in the city trade are often intermingled with residences.

The pattern of urban rents is complicated by zoning restrictions, historical accident, and the revolution in local transportation brought about by the motor vehicle. Zoning restrictions are an administrative allocation of total space within a district for various uses, and therefore often limit the use of specific locations for purposes which would bring higher rentals. In many business districts there may be found single-family residences still occupied despite offers from business to buy the properties. Factory and residential districts are often mixed. Residential districts in rapidly growing cities often engulf farms.

QUASI RENT

The discussion of the rent of land alone has been undertaken thus far because land and permanent improvements to land represent sources of productive services which are fixed in supply under *long-run* conditions. In the *short run,* however, the sources of many other types of productive services may be considered to be fixed in supply and the prices for such services may be determined in a manner similar to that for rent. Prices of services in fixed supply other than the services of land are termed *quasi rents* because the rent analysis is also applicable to them.

Capital facilities already produced represent probably the largest source of services whose prices in the short run are considered quasi rents. Once factories, machines, office buildings, or other types of capital facilities have been built, their services remain available during the period of their use-

fulness—whether or not the prices paid for their services reflect the going return on the cost of the capital invested. That is why the discussion of the short-run equilibrium of the firm indicated that the firm would produce so long as it earned any quasi rent, i.e., so long as it covered average variable cost.

When an investment in capital facilities is first made, it is always expected that it will yield at least the market rate of interest on the capital invested; and if commodities of a similar type do not yield at least such returns, no more of them will be built. That is why the long-run costs of the firm include depreciation and interest on capital; in the long run the firm must cover average total cost, not just variable cost. When investment is made in buildings or machines, however, it is committed almost irrevocably. For a considerable period of time the capital facility will render services, and it will do so regardless of its earnings. In short, it becomes for a time a source of a fixed supply of productive services of a certain type.

The price for the services of a machine or building already in existence depends upon the demand for its services, just as rent depends upon the demand for land as reflected in the schedule of the value of the marginal product of land. Since machines and buildings are likely to be built for specialized purposes, the possibility of varying the proportions of the other services used in combination with them is small. In many cases it is therefore impossible to determine a price for the services of a capital good which is independent of the price for the other factors used with it. For example, if a machine were designed to be run by one man, the combination of labor with machine services could not be readily varied. Although the value product of man and machine working together is determinable, the price received for the machine services can be calculated only by deducting from the total product the cost of those services which can be independently priced. If man and machine add $10 per day to the value of materials and power furnished them, and if the man's wage is $5 per day, then the services of the machine must yield $5 per day. This return, of course, may be above or below the current rate of return on the capital invested in the machine, and has—at the time—no relationship to the supply of such machines.

There is a minimum price, however, below which it will not pay to use any durable capital good already in existence. If the capital good is suited for one use only, and has no value in any alternative use, then it would pay to accept whatever price can be obtained for its services, since the alternative is to obtain nothing at all. However, most such capital goods have some alternative use either in the present or the future. A machine, for

example, may be sold for scrap, the price of which would set at least a minimum price for the machine's services; at any price below this it would pay to scrap the machine. A higher minimum price than the scrap value is usually set, nevertheless, because machines can be used either in the present or the future. If the machine depreciates through use, the price received for current use must at least equal the price which the owner expects could be received in the future (less the storage, maintenance, and interest costs carried from the present to the future period). Price expectations, however, are important; the owner of a machine who has some degree of control over the price of his product may not accept a current price which he considers too low for fear of spoiling his market, because he expects a higher price in the future.

There is also the possibility of using the capital good for another than its original purpose both in the present and in the future. A building originally constructed for use as a factory might be used as a warehouse, or a store might be used for an office. All of these considerations—present and future prices in the same and alternative uses—usually result in the owner's placing a minimum reservation price upon the services of any durable capital good in fixed supply. Because reservation prices set on durable capital goods are sometimes higher than current prices, the capital goods are held idle, even though some income could be received for their use.

Quasi rents apply not only to durable capital goods but also to human services. It has been emphasized that quasi rents are the prices for services in fixed supply. If the services of a person are virtually unique, then they become for some time—for purposes of analysis—services in fixed supply. Thus, motion-picture stars command enormous prices as a source in fixed supply of productive services which, when translated into sound track and celluloid, have great box-office attraction. Similarly, the presidents of large corporations who have convinced their stockholders that they are uniquely responsible for the success of the company and who are in fixed supply command high quasi rents in the form of high salaries. During World War II many types of skilled workers had special skills which were in fixed supply and which commanded high wages.

Quasi rents are not always high, however. For many years the supply of vaudeville performers' services was virtually fixed, but since there was little demand for vaudeville entertainment, performers' earnings were low. It is the fact of fixed supply which gives rise to quasi rents; supply being fixed, demand is the important determinant of the amount of quasi rent. Of course, in the long run the supply of factors yielding quasi rents can be changed, and the prices for their services reflect long-run supply and demand. Thus

the supply of capital facilities or skilled workers is adjusted to demand in the long run, whereas in the short run the supply is fixed.

RENT AND COST

Under what conditions is rent a cost? From the point of view of an individual firm which requires the services of land, the firm must either pay a rent or pay a purchase price for the land which reflects the value of its marginal product. Rent is clearly a cost to a firm which must pay it or do without the services of land. Assuming that the competition for the space is among firms operating under purely competitive conditions, the firm must match the rent which another firm might earn by using the same space, and in this way the space is utilized to the best advantage from the standpoint of the market demand. If the user owns space, he must earn the competitive rent or else it would pay him to rent it out or sell to someone else. If one firm can conduct its business much more advantageously in a certain space than any rival in the same line, then by paying a rent equivalent to that offered by its rivals the firm will yield to the entrepreneur a "rent" of superior enterprise, or profit, which other firms do not enjoy. This rent of enterprise can be considered to be in the class of quasi rents for rare ability, and would represent a surplus.

This reasoning may be applied in the same way to the returns from land devoted to producing different crops. A given area of land devoted to oranges might yield $25 per acre; and its next most advantageous alternative use might be for truck gardening yielding $20 per acre. The excess return over the next highest alternative use, $5, represents a rent to orange growing; it is a surplus in that it is a return in excess of the cost necessary to obtain the supply of land for orange growing. From the individual fruit grower, however, a landowner might be able to obtain $25 per acre for the use of the land, since growers would compete with each other for the privilege of using the land worth $25 per acre to them. Thus, while $25 per acre may represent a cost to the individual grower, from the standpoint of assuring the use of the land for orange growing, $5 per acre is surplus.

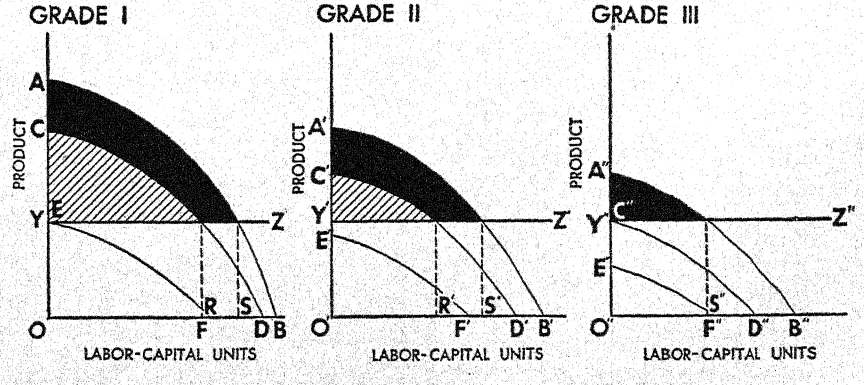
To sum up, all rent or quasi rent can be considered to be a surplus, i.e., *a price not necessary to assure supply from the standpoint of society as a whole.* However, rent is necessary to assure the use of productive agents in a *particular industry,* for it represents an amount equal to the earnings of the agent in its next highest alternative employment in another industry. The difference between the earnings of an agent in one industry and the

(lower) earnings in its highest alternative employment can be considered to be a surplus in the sense that it has no bearing upon supply. Finally, so far as an individual firm is concerned, rent equal to the price of the service of the agent in a competitor's hands must be considered to be a part of the cost of the firm, since this price must be paid for the agent or earned by it to retain the service in the firm's use.

In considering the relation of rent to supply, it should be noted that rent performs the function of ensuring the allocation of productive agents to those uses which consumer demand values most. This function of rents in allocating resources can be performed in any kind of economy, of course, whether or not there is private ownership of land and durable agents of production. By ensuring that land and durable agents of production receive the highest possible rents under competitive conditions, an economy makes the use of land and durable agents conform with the demands of the public. Even a socialist state holding title to all land and durable agents of production must follow this principle if it aims to ensure production in conformity with the private demands of its citizens.

The nature of rent is illustrated graphically in Chart 18. Each diagram represents a certain amount of a given grade of land in its most productive use. On all the diagrams, the vertical axis measures the amount of product; the horizontal axis, the number of units of a given combination of capital and labor applied to the fixed amount of land. The sloping curves, *AB*, *CD*, and *EF* represent the values of the marginal products as more units of labor and capital in fixed combination are applied to a given amount of land of a particular grade. The horizontal lines *YZ* represent the cost per unit of product for the units of labor and capital.

Chart 18 RENT ON THREE GRADES OF LAND



The curves *AB*, *CD*, and *EF* all slope downward and to the right because of the operation of the law of diminishing returns. As more units of labor and capital are applied to a given amount of land, the addition to the total product resulting from the addition of the labor and capital falls off. The law of diminishing returns, therefore, accounts for the *slope* of the curves. Since the curves represent the values of the marginal products of the labor-capital units, their *level* depends upon the price of the product.

Suppose, for example, that the price of the product is so low that the values of the marginal products are represented by *EF*, *E'F'* and *E''F''* on the three grades of land. The labor and capital cost is *OY* per unit of product. Since the value of the marginal product nowhere exceeds its cost, the three grades of land will lie idle and none will yield any rent. Some grades of land may have some value but only because they are expected at some future time to yield rent.

If the price of the product rises, the curves showing the values of the marginal products will rise also. When the curves *CD*, *C'D'* and *C''D''* represent the values of the marginal products, an entirely different situation than before results. It pays to use labor and capital on grade I land in an amount represented by *OR*, and on grade II land in an amount represented by *O'R'* since the value of the marginal product exceeds the cost of the factors up to points *R* and *R'*. On grade III land, however, *C''D''* lies below *Y''Z''* at all points—the value of the marginal product nowhere exceeds its cost—so that grade III land is not used.

Now examine the effect on rent of an increase in the price of the product. It must be remembered that the area lying under the marginal product curve represents the total product, since the marginal product curve represents all *additions* to the total product, i.e., from no product upward. Obviously, the sum of all additions, from *O* upward to any point, must be the total product. The total cost, likewise, is measured by the area under the curve representing additions to cost, *YZ*, *Y'Z'*, etc. The shaded areas for grade I and grade II land are, therefore, returns in excess of cost and are—under the conditions assumed—rents. Thus when prices are high enough, some land yields rent, and the best grades yield more rent than the lower grades. The poorest grades of land, represented by grade III, yield no rent. It is true, therefore, that prices are not high because rents are high, considering the economy as a whole, but that rents (and land values) are high because prices are high.

It may be shown finally what happens in overcrowded countries. When the values of the marginal products rise to the levels represented by *AB*,

$A'B'$, and $A''B''$ on grades I, II, and III land respectively, even grade III, the lowest grade of land, begins to yield rent (black-colored area), and the rent on grade I and grade II land increases (by the amount shown in the black-colored areas).

If the three diagrams are considered to represent urban land, they illustrate why tall buildings are built in downtown areas. If AB, $A'B'$, and $A''B''$ represent the values of the marginal products from applying labor and capital to urban land, it would be profitable to apply OS units of labor and capital to the best location, $O'S'$ to grade II locations and $O''S''$ to grade III locations. Since the level of the curve showing the values of the marginal product is highest for the best location, the amount of labor and capital (in the form of height of building, number of operating and maintenance workers, etc.) which would be used with a given amount of land in this location is much greater than for other less productive locations, i.e., OS is greater than $O'S'$, which in turn is greater than $O''S''$.

CHAPTER 21

Profits

Profits mean for most people, including businessmen, the return remaining from the receipts of a business after expenses have been subtracted. This conception of profits produces some interesting results which can be simply illustrated. Suppose that two men, Jones and Smith, enter the retail grocery business, each selling over a year's time the same volume of identical goods. Suppose furthermore that Jones puts up his own money out of his savings, whereas Smith borrows his money from a bank, giving a mortgage on his house; and that Jones buys his store property outright, whereas Smith rents his store. Smith was formerly employed as a store manager at $2,400 per year, while Jones had formerly owned another store. At the end of the year of operation, each draws up a statement of profit and loss as follows:

Jones' Grocery

Profit and Loss Statement

Net sales		$100,000
Cost of goods sold:		
Purchases	$95,000	
Less inventory	5,000	
		90,000
Gross profit on sales		$10,000
Operating expenses:		
Wages	$2,400	
Insurance	50	
Depreciation	250	
General expense	100	
Total expense		2,800
Net profit		**$7,200**

Smith's Grocery

Profit and Loss Statement

Net sales		$100,000
Cost of goods sold:		
Purchases	$95,000	
Less inventory	5,000	
		90,000
Gross profit on sales		$10,000
Operating expenses:		
Rent	$1,200	
Wages	2,400	
Salary	2,400	
Insurance	50	
Depreciation	250	
General expense	100	
Total oper. exp.	$6,400	
Interest on mtge.	600	
Total expense		7,000
Net profit		**$3,000**

Jones' store shows a net profit almost two and one-half times that of Smith's store, yet the goods sold, their cost, and the volume of business were identical. Obviously, the difference in net profit is simply due to differences in costs or, more accurately, in what are believed to be costs. Jones pays no rent because he owns his store, yet he could rent his store for at least $1,200 (since Smith must pay this amount). Jones charges no salary or interest against the business, but Smith includes a salary equal to what he would have earned as a store manager; furthermore, since Smith has borrowed capital he charges the interest which he has paid. The difference in net profits, therefore, results from Jones' considering *explicit* costs only.

These differences between the true economic position of the firm and the position as revealed by its accounting books make it necessary to distinguish between "business profits" and "pure profits." Business profits are simply the difference between receipts and explicit costs (these are properly shown on Jones' statement of profit and loss). Pure profits, on the other hand, consist of the difference between receipts and implicit as well as explicit costs. Pure profits reflect the true economic position of the firm. Thus, the opportunity cost of the manager (Smith's $2,400 salary) must be included as well as the opportunity cost of capital (what Jones would have

earned by investing his money on the best alternative opportunity). Rent must be included at a figure equal to the rent which would be received from the best alternative use of the land or space. Defining pure profits in this way, it is possible that Jones could show a net business profit of $7,200 but a smaller or even negative pure profit. He might, for example, hire a storekeeper for $2,400 a year and draw from his store—with the capital invested in it—$4,800. If he himself were to obtain a job paying $4,000 per year, he would gain by $1,600 per year ($8,800 instead of $7,200), assuming that his manager runs the store as efficiently as he did. It follows that "pure profits" must be understood to be either positive or negative (the latter being losses). Thus, there is a world of difference between what most people consider to be profits and what pure profits are.

PROFITS IN LONG-RUN STATIC EQUILIBRIUM

Perhaps the nature and extent of pure profits can best be seen by considering a set of conditions in which profits cannot exist and then by introducing certain changes into these conditions to show how profits can arise. In this manner it is possible to isolate for separate study those conditions in the actual economy which are responsible for pure profits.

Assume that pure competition exists throughout the economy, that the state of technology, the supplies of productive services and their proportional combination, and the tastes and incomes of the people remain unchanged. Under these conditions the price of each product is equal to its average cost; productive services are combined in such a way that the ratio of their prices is equal to the ratio of their marginal products, and the lowest cost combination is therefore reached; finally, the price of each productive service is equal to the value of its marginal product. Payments to each type of productive service during any given period will be the price of the productive service multiplied by the number of units supplied; these payments, constituting the income of the economy, are in total equal to the value of all goods produced. This is the only set of relationships which, under the circumstances, is consistent with an equilibrium condition for all firms and for the economy as a whole.

Under such conditions, all economic life would consist of a repetitive process. The operation of each firm is reduced to a routine function which can be performed by people working for a stated salary, which amount will be regularly earned by both the firm and the recipient and will be considered a part of the firm's cost. So long as the firm can cover this and other costs it would continue indefinitely to produce. Each firm's costs would depend

upon the prices of the productive services it uses, and these would be dependent upon the values of their marginal products, which must be equal in all lines.

Thus, if price is equal to average cost for each firm no firm would make a pure profit; and the fact that the services of the manager of every firm can be hired for a stated salary makes it possible to treat the manager as a supplier of (skilled) labor services. There are no functions performed except those of labor, capital, and land, and there would be no prices for productive services other than wages, interest, and rent. In these circumstances, termed a *static long-run equilibrium,* there are no pure profits either for firms or as a payment for productive services. In the static equilibrium there could be no uncertainty as to the outcome of a given firm's business because all economic life under the conditions assumed is repetitive. Even such disturbances as fires, robberies, and accidents could be withstood in a repetitive type of economy without loss to any firm, by means of insurance with premiums calculated exactly to offset losses in the economy as a whole.

PROFITS IN A DYNAMIC COMPETITIVE ECONOMY

Suppose now that some element of change is introduced into the completely repetitive economy, the results of which cannot be accurately predicted on the basis of past experience. For example, remove the assumption that tastes are constant. If people constantly change their tastes, demands for some goods will rise, and for others will fall: no one knows which goods will be affected by increases in demand and which by reductions. There is now an element of uncertainty in the economy which puts a premium on good guessing as to what the resulting shifts in prices will be. Again, to bring out another point of importance, suppose that all firms consist of co-operatives made up of the suppliers of the productive services. Under the new conditions of uncertainty, the suppliers of productive services will attempt to predict what kinds and amounts of goods are most profitable to produce. They will finally undertake production according to a plan they agree on. This may involve committing capital in the form of goods and machinery suitable primarily for carrying out the plan agreed upon. There remains no agreement, however, on what prices will be paid for the productive services required.

Production is carried on according to plan for a period of time. If workers and other suppliers of services can depend on immediate alternative uses, they may demand and receive prices for their services equivalent to what they might obtain elsewhere. On the other hand, suppliers of the more

specific sources of services, such as machinery, may have no significant alternative market for their services; therefore, they must wait for the sale of the goods produced over a considerable period of time before they can recover their outlays. If, over a period of time, the plan is successful and the goods produced can be sold at prices in excess of costs (calculated on an opportunity cost basis), then the enterprise will make a pure profit. If the estimates upon which the plan was based turn out to be too optimistic, prices will not cover costs and a loss will result. Owners of sources of productive services which can be shifted immediately to other uses may insist upon prices equal to those obtainable in other uses, or else withdraw to the other uses. But the owners of sources of productive services having *specific* uses in the type of production undertaken, and consequently little alternative use, cannot obtain their prices elsewhere and must take the chance of obtaining a profit or taking a loss. Thus, under a co-operative production plan where uncertainty is present, the owners of capital goods are likely to obtain profits or to bear losses.

Because of the difficulties of operating under a co-operative production plan—mainly because most suppliers of productive services wish to substitute a certain price for an uncertain one—a system of contracts is needed in which one person or group assumes the right to the profit or the responsibility for the loss. Historically, this person or group has usually supplied capital, probably because he or they owned sources of services of the most specific type. However, many suppliers of capital have in the past and do presently choose not to bear uncertainty, so that a new kind of productive service has appeared: that of the entrepreneur, who undertakes the combination of productive services under conditions of uncertainty for the return, whether profit or loss. Only if the resources of the entrepreneur are largely wiped out do the suppliers of the other productive services stand to bear the uncertainties of enterprise; and these uncertainties can of course be greatly reduced by requiring the entrepreneur to furnish security that his contract will be carried out. Thus, the risk of losses arising from uncertainty can be shifted to the shoulders of some one person or group. Since the desire for security is strong, such assignment of risk (and of profits and losses) has widely characterized economic activity for centuries.

THE ENTREPRENEUR

It is under conditions of uncertainty, therefore, that there arises an entrepreneurial service—contracting to pay certain prices to the suppliers of other productive services to produce a product which may sell for more

or less than its cost of production. Since the entrepreneur assumes the chances of loss of all the owners of productive agents with whom he contracts to pay definite prices, it is natural that he should be given control over the productive services needed to meet his obligations. The function of the entrepreneur, consequently, consists of two parts: (1) assumption of chances of loss and rights to profits, and (2) ultimate decision as to how productive services are to be used.

In relatively small business concerns the entrepreneur usually can be identified as a person, and he will be found to perform the two functions stated: he may invest his own capital in his business, so that all his property is subject to forfeiture if he fails; he may personally direct his business and receive all the profits it earns. In larger concerns, especially in large corporations, the function of the entrepreneur may be shared by a number of people. Some stockholders (usually holders of preferred stock) who own an interest in the corporation may choose to remain "contingent entrepreneurs," i.e., they contract away their right to any control over the corporation except when their promised dividends are not paid. Another class of stockholders (common stockholders) may accept a claim to only part of the profits in return for the right to control the corporation and to receive a major part of any large profits which their nominal or actual control brings to the corporation. These stockholders holding the legal right to control may also choose to delegate it to a board of directors or to the president of the company.

As the function of entrepreneurship has been dispersed among many people in a modern corporation, the "entrepreneur" cannot be identified as a person. The function of enterprise must, nevertheless, be performed, and the price paid for this function consists of profits or losses. The function of enterprise is exercised by whatever group of people at any one time together hold the right to control the corporation and the disposition of its profits.

It has already been shown that the measure of pure profit to a single entrepreneur is not the whole amount that he earns, but rather the return over and above what he might earn in his best alternative employment. It follows, then, that returns from performing the services of enterprise that are lower than earnings in the best alternative employment are negative pure profits, or losses. This same principle can be extended to the many persons in a large corporation who share the function of enterprise. For such persons who have invested capital in the corporation, pure profits are received only when the return on their investment exceeds what they might have earned in investments carrying the lowest chance of loss. And if they

spend time performing their function, they receive pure profits only when they earn more from it than they might have earned in their best paying alternative employment. It is obvious that the total amount of pure profits defined in this manner is likely to be very much less than the "profits" which people are considered to receive from dividends.

PURE PROFITS IN THE MODERN ECONOMY

In an economy where pure competition prevails and where the only uncertainties are those which influence relative prices, such as shifts in demand among goods (which was introduced in the preceding discussion) or improvements in the techniques of production, it is likely that some firms will show pure profits and some losses. Assuming further that there is a normal distribution of ability among entrepreneurs, profits and losses of particular firms would tend to cancel out, and thus profits would probably not exist as a total in the income of the entire economy. Under the more numerous conditions giving rise to uncertainty in the modern economy, however, it is possible for pure profits or losses to occur not only for individual firms but for the economy as a whole.

One of the most important conditions giving rise to pure profits or losses in the modern economy is general movements in either the level of prices or in the level of national output. The individual firm has no control over either. When the level of national output rises, and especially when it does so to the accompaniment of rising prices, most firms make pure profits. This happens because the prices of productive services are not immediately increased to correspond with changes in the values of their marginal products. The rate of interest on capital may not be adjusted immediately because certain rates possibly have become customary; moreover, especially under modern conditions interest rates may be determined by the government credit policy to a greater extent than by market influences. As has been indicated, wage rates may not be adjusted to the increased value of labor services for some time, especially if there exists some unemployment throughout the economy. Moreover, as the volume of production increases the amount of overhead costs of business remains unchanged for some time, with a consequent favorable effect upon profits. Profits earned by firms during a general upward movement of business are not immediately offset even in competitive industries, because a considerable period of time must elapse before new firms can increase the supply of goods and bid up the prices of productive services to the new values of their marginal products.

A downward movement of the level of prices and of national output occurs in depressions. Prices of goods generally fall faster than the prices of productive services, output declines greatly, and pure profits become negative (i.e., losses). The changes in *business profits* during periods of prosperity and depression are likely, however, to be greatly in excess of the changes in *pure profits*. During the upswing, for instance, business profits are magnified by the practice of recording interest and depreciation on the basis of contract or customary rates; whereas in calculating pure profits, both interest and depreciation would be calculated on an opportunity-cost basis and would therefore be adjusted upward as interest rates in the market and the replacement cost of capital goods increased. At a time of rising prices, business profits include the appreciation of the value of goods sold from a lower cost inventory, whereas in calculating pure profits inventory would be valued at current or market prices. Business profits during a depression are probably lower than pure profits because contract costs incurred during a period of prosperity are carried over into a lower price period, and because high-cost capital facilities and inventory are "written down" to reflect current lower market prices. In tracing the movement of pure profits, therefore, financial statements [1] do not offer a reliable guide.

A real issue is whether pure profits during periods of business prosperity or losses during periods of business depression are properly attributable to the good or poor performance of the functions of entrepreneurship. Most of the profits or losses which appear during such periods result from general movements in the economy beyond the control of the individual firm. It follows that these unusually large profits or losses are mainly unrelated to the functions of enterprise as outlined here, and in reality must be considered mostly as windfall gains or losses rather than prices paid for the performance of necessary services.

Another condition giving rise to pure profits is to be found in the frictions or imperfections of competition, which may result in profits or losses to individual firms and perhaps even for the economy as a whole. A manager may be paid either more or less than he is worth in terms of his alternative opportunities. The same is true of labor services generally. Absence of complete knowledge of alternative opportunities or of sufficient mobility may keep the prices of such services from equaling their prices in alternative uses. Imperfections of competition may likewise prevent the prices of other productive services or of goods from reflecting the going market levels. Loan rates in country areas are almost always higher than in urban areas for this reason. Some frictions may actually be created by law, as when

[1] Such financial statements are described in Chap. 26.

tariffs are raised to shut out foreign competition, or when wages are raised by minimum-wage legislation.

A further condition giving rise to pure profits is monopoly power. Monopoly power may be created by the individual firm through product differentiation or by the exercise of its dominant size in an industry, or by mutual restraint in an industry composed of a few sellers, or by any one of the many devices which give to producers some power over the prices of their products. Monopoly power may also result from the operation of laws which restrict competition, such as the patent and copyright laws, and public franchises granted to railroads and utility companies. This power obviously gives a positive basis for earning pure profits to any firm possessing it. The existence of monopoly power in the economy tends to create a positive amount of pure profits for the economy as a whole, since it permits many of its possessors to exact returns in excess of those reflecting opportunity costs.

Still another source of pure profits can be classified generally as "innovations," which may include inventions which improve productive methods, or new technology in the methods of management and labor, or the development of entirely new products. Even under purely competitive conditions, innovations are likely to yield pure profits to the innovators. Some innovations are patented and so ensure pure profits (within limits of change) by legal monopoly; others result in pure profits to innovators until competition is made active, i.e., until enough competitors have had time to imitate the innovator. It therefore appears that innovations lead to pure profits for individuals, and almost certainly for the economy as a whole.

Are Pure Profits Necessary?

Pure profits, then, result from (1) the existence of uncertainties, (2) general movements in prices and the level of national output, (3) imperfections of competition, (4) monopoly, and (5) innovations. It might be said that the economy would fare better without the pure profits resulting from monopoly and from the general movements of output and prices. Under conditions of uncertainty, however, profits serve in some degree to secure the allocation of resources to those lines most preferred in terms of consumer demand. Losses in one line over a long period of time will result in a flow of resources to more profitable lines. This function of profits depends upon the assumption that profits arise because anticipations of consumer demand in certain lines have proved to be incorrect; if, however, profits result from monopoly (except possibly such socially

desirable monopolies as patents), they cease to perform this economic function.

Granted that profits perform a necessary function in guiding resources into lines preferred by consumer demand, the question may still be asked: Is it necessary for profits to be earned by the economy as a whole? Is it essential to the well-being and continuance of the economy that its total profit-and-loss statement—after all profits and losses of all enterprises are entered—show a positive figure? Some economists have argued that if uncertainty is the only cause of profits, so long as some individuals make profits there is probably no necessity for pure profits to be earned by the economy as a whole. But the fact remains that profits are also linked up with progress. They are a reward for successful innovation. Indeed, there is some question whether in an economy where innovations are constantly occurring, pure profits would ever disappear completely. Although an innovation may be made only on the questionable chance that the innovator will enjoy pure profits, innovations are made in such great numbers that the pure profits earned by the many innovators are not clearly offset by pure losses elsewhere in the economy. This is especially likely because knowledge and mobility are not perfect.

As a motive for investment, the chance of earning a pure profit serves as one of the mainsprings that drives the economic mechanism. People will more willingly risk their savings—and businessmen their accumulated capital —in new processes, new machinery, or new products when the chance of large profits is present. Pure profits, therefore, serve not only to allocate resources among existing lines of business but also to encourage the development of new production, to stimulate new investment, and to propel a constantly rising level of national output and national income.

Profit and Risk

Profits arise because production must be carried on under conditions of uncertainty and change. Nevertheless, there is a strong desire to substitute certainty for uncertainty and security for insecurity in the conduct of business and personal affairs throughout the entire economy.

One method of dealing with uncertainty is to prevent it from arising. Risk of loss of buildings by fire can be reduced, if not eliminated, by using fireproof building materials. Loss by theft can be prevented by employing honest workers, by making property burglarproof, and by securing adequate police protection. The individual worker can reduce the risk of losing his job by steady and outstanding performance; and workers generally

may reduce the chances of unemployment by sponsoring governmental policies which are aimed at full employment.

In addition to eliminating the chance of loss at the outset, risk can also be prevented by improving the methods of prediction of the course of events. Mainly this is done by specialization of knowledge. A businessman can predict the course of his business more accurately than an outsider; an engineer can predict the load a bridge will bear better than the motorist who crosses it; and a labor leader can generally predict the chance of success of a demand for higher wages better than the rank-and-file union worker.

Another method of preventing risk is by combination of several risks. A manufacturer who diversifies his production may compensate for the loss of the market for one product by what he gains from another. Manufacturing concerns like General Foods have attempted to gain by such combinations, as have retailers such as department stores.

If risks cannot be eliminated by any of these methods, the individual may nevertheless remove the burden of risk from himself by placing it on others or by shifting it. As a part of the process, the risk itself may be reduced through specialization or combination by those to whom it is shifted. The classic method of shifting risk is by insurance. Insurance is based on a calculation of the average probability in the occurrence of events, it being determined by an analysis of a large number of independent cases of the same class. If a coin is tossed a great enough number of times, heads will turn up equally as often as tails. On any particular toss, however, the result cannot be predicted. The same principle can be employed to establish the probability of loss from accidents, fire, theft, or termination of income because of premature death. If on the average it is found that the fire losses of frame farm buildings amount to $2 per $1,000 of insured value, then a fire insurance company can determine the premium necessary per $1,000 of insured value to pay off policyholders and return a profit to the company. Since the loss of buildings by fire is a large risk for the farmer, he is willing to pay a few dollars per year in premiums to relieve himself of bearing the risk. For the insured, therefore, insurance is nothing more than the substitution of a small but certain cost for the risk of a large loss. In life insurance the same principle applies, for it is not insurance against death—which is certain—but against a total loss of income because of death.

Beyond buying insurance it is still possible to shift risk by other means. It has already been shown that suppliers of materials and of labor services succeed in shifting most of the chance of loss on their services to the entrepreneur, since they contract to receive a certain return regardless of the fluctuations in the market for the product which they aid in producing.

Businessmen can sometimes reduce the chance of loss by taking orders in advance for their products, and also by contracting in advance for the materials needed to fill the orders. For certain standard commodities which can be stored, the market is so well organized that a specialized form of risk-shifting called hedging is possible. To "hedge" is to take an identical position on both sides of the market so that price increases or decreases cannot effect a loss (or gain), as when a person bets an equal amount on both teams in a football game. It happens that cotton or wool textile manufacturers, flour millers, and many other businessmen must of necessity carry inventories of their raw material. If they wish to confine themselves strictly to their manufacturing and distributing operations without risking inventory losses because of price fluctuations in raw materials, they can shift the risks by hedging.

A third general method of dealing with risks is by setting aside profits when they are earned against subsequent losses. For many risks can neither be insured against nor shifted economically to others. Farmers may be unable to make a crop because of drought or pests (although crop insurance may in some cases be possible); workers may lose their jobs and remain unemployed beyond the period of unemployment benefits; and businessmen may lose heavily because of loss of markets, the advent of improvements which render their plants obsolete, or for other reasons. These risks must be borne but their incidence can be mitigated by making compensatory plans or setting aside reserves. Wage earners "save for a rainy day"; business enterprises set up reserves for contingencies; people whose jobs are uncertain keep up their "contacts."

It is apparent that uncertainty imposes an economic burden upon society. When economic resources—whether the services of people or of things—are devoted to prevention or shifting of risks or to preparation for meeting them, these resources cannot be used for other productive purposes. Therefore, it may be said that a reduction of risk in economic processes frees resources for other uses and contributes to a higher standard of life.

In a free-enterprise economy, the reduction of the uncertainties faced by the entrepreneur can have a considerable bearing upon the level of employment, since the function of the entrepreneur actually consists of combining the services of productive agents under conditions of uncertainty. The willingness of businessmen to contract for productive services depends ultimately upon their estimates of the chances of making profits; however, businessmen are seldom certain to make pure profits but instead can assume only some sort of average probability of making a profit. Changes in the market or economy as a whole which reduce the probability of making a

profit are likely to be depressing to business. These changes may occur in the prices of productive services, in the prices of the products produced, or in the ability of businessmen to retain earned profits. Thus there is a constant appeal by businessmen for "stability," and a repeated declaration that "confidence" of businessmen in the future is necessary for prosperity.

Although the expressed need for "stability" and "confidence" is admittedly often the excuse for monopolistic arrangements, there is a more fundamental reason for the business world's objection to the changes wrought by many innovations. Businessmen faced with a multitude of uncertainties object to the imposition of additional uncertainty by certain actions of government when these actions are likely to affect their profits adversely. Changes in tax laws (especially if upward), the adoption of new programs such as deficit spending which may require heavier taxation in the future, and the adoption of many social reforms—whether favorable or unfavorable to society as a whole—represent new uncertainties for businessmen, and probably exercise for a time some depressing effect upon their willingness to make new commitments. The usual political conservatism of businessmen is to be expected in the light of this fact.

SUMMARY—PRICES OF PRODUCTIVE SERVICES

The bulk of economic activity is directed in the last analysis toward satisfying consumer demand.[2] In this economic system, consumers express their preferences by purchasing a myriad of goods and services in the stores, shops, and market places of the nation. These preferences govern the production of goods and services and, therefore, the allocation of labor, capital, land, and other economic resources. The daily purchases of the nation's consumers and the practices of sellers produce a constantly changing pattern of demand and supply for goods in all the markets of the country. On the basis of market prices and the prices which are expected to prevail in the future, producers of commodities and services undertake to supply the kinds and quantities of goods and services which appear likely to maximize their profits, or at worst minimize their losses. In undertaking this productive work, enterprisers demand, directly or indirectly, productive services of all kinds—the services of labor, capital, and natural resources. These demands of enterprisers are always conditioned by the fact that the law

[2] This refers to the individual- or family-consuming unit. The remainder of economic activity is devoted to satisfying social needs as represented by governmental activity. Normally most governmental activities directly or indirectly minister to the wants of consumers, e.g., the maintenance of order and the protection of individual and property rights are indispensable to the operation of a private enterprise economy.

of diminishing returns is at work, and that it will not pay to hire or purchase more than certain amounts of each type of productive service alone or in combination with others. Thus, each enterpriser organizes his productive unit, combines labor, capital, and other resources, in such a way as to enable him to supply those who want his goods and so earn a profit.

The aggregate of goods and services thus supplied comprises the bulk of gross national output; and the aggregate of wages, salaries, interest, and rent paid out comprises the bulk of national income. The sizes of these aggregates, the composition of the output, and the allocation of economic resources are all governed by producers' expectations of costs and receipts, or prices of the things they sell and the prices of the things they buy or hire. The volume of sales, and the prices at which sales can be made, directly or indirectly determine the demands of enterprisers for productive services of all kinds. The demand for and the supply of each type of productive service determine the price paid for it. The prices of productive services are the incomes for personal services and for the owners of capital and natural resources, and this income comprises total national income.

The amount and distribution of income, in turn, profoundly affects the future level of production, since the amount of income which consumers as a whole devote to spending is normally accompanied by an amount (and types) of goods which will sell, on the average, at the prices expected. And the savings out of income will normally be accompanied by an amount of capital goods production (investment) to absorb these savings. The relationships between spending and the production of consumer goods and services, and between saving and investment, have been discussed earlier.[3] It is sufficient to repeat here that expected prices in the near future govern the short-term output of consumer goods and services (and the income paid out in their production), while expected prices in the more distant future govern long-term investment. The composition of consumer goods output and the composition or types of long-term investment depend upon expected relative prices and consumer expenditures. Production undertaken on the basis of these expectations comprises the demand of enterprise for productive services, and the prices paid for these services are the incomes of productive agents.

In a capitalistic society, the directing force for most economic activity is the price system, and the balancing force is profit. It is obvious that the maintenance of a satisfactory level of national income depends upon the proper co-ordination of the prices and output of consumer goods, the prices and output of the factors of production, and the allocation of income be-

[3] See Chap. 5.

tween saving and spending. When, because of monopolistic practices, wars, monetary disturbances, political upheavals, or similar influences, the nicely adjusted direction and balance of economic activity by the price system and the profit motive are impaired, intervention by government is often sought in order to restore or modify the price and profit system in the hope of attaining an adequate level of national income and an appropriate allocation of economic resources.

CHAPTER 22

The Price Structure and Price Movements

The discussion in the preceding chapters has shown that the relationships among prices influence consumer purchases, the level and composition of national output, and the allocation of resources.

In varying degrees each price is dependent for its formation upon all other prices in the system; but the price of any particular commodity or service is immediately and directly related only to the prices of a group of competitive or complementary commodities. The prices of industrial fuels, for example, are so closely related to each other that many power plants are equipped to burn either oil or coal, and usually several grades of both; but the prices of industrial fuels are only remotely related to the price of bananas. The prices of beef, hides, and tallow, which are complementary products, are affected jointly, if not in the same way, by their common source of supply, but the relationship of the prices for all three to the rates of interest on long-term government bonds is negligible. The prices of hand-tailored or custom-made clothing are related to the prices of factory-made clothing, but both are little affected by the prices for airplanes or air-line transportation.

It is customary to consider the relations among prices prevailing at any one time as the *structure* of prices. To explain how the market, short-run, or long-run prices of goods and services are determined, as has been done in the preceding chapters, is in a sense to explain how the structure of

prices is formed. A structure of prices is, then, a temporarily static grouping of all prices in the economy from which changes are measured and movements charted. The study of the structure of prices is essentially a study of relative values. A change in the price structure may result merely from an increase or a decrease in the price of one good or several goods, relative to the prices of other goods or the prices of productive services.

Changes in the structure of prices must be distinguished from uniform proportional movements of all prices in the same direction. If all prices moved upward or downward in proportion there would not be any changes in relative prices. But if prices moved unevenly up and down—as they usually do—the economy experiences a change in the structure of prices. This chapter is concerned with price movements which occur over a period of time—from one point in time to another—most of which involve changes in the structure of prices.

When the economy is confronted by new conditions, movements over a period in some prices—relative to others—are means by which it adapts itself. This is evident when it is seen that many such movements are essential to the maintenance or increase in national output. If, for example, during a period of rapidly improving technology wages do not rise relative to the prices of commodities affected, income and therefore consumption may not be able to absorb the increased amounts of commodities produced by the new technology. If rayon prices had not fallen relative to other prices when drastic cost reductions were made in producing rayon, production soon would have outrun consumer purchases. If the price of wheat in the European market had not fallen when American production from the newly opened lands in the West became available, the United States would have been settled more slowly; by the same supposition, British farmers would have had more prosperous times but British workers would not have been so well fed. Although changes in relative prices are not always the result of economic progress, they are a necessary concomitant to any improvement in the allocation of resources or production of goods.

The extent of movements in relative prices is remarkable. During the period 1896-1913, when the level of wholesale commodity prices rose at the rate of 2.3 per cent per year, some prices—notably interest rates—went down; some prices, such as that of sugar, changed hardly at all, and others rose at the rate of 3 or 4 per cent per year. This can be seen in Table 47.

A similar pattern of divergent relative movements will be found in any period of a few years. During World War II, of course, the spread among prices was widened by the extreme pressure of war-swollen demand on the one hand, and dislocations of production which unevenly affected costs on

Table 47 **AVERAGE ANNUAL RATES OF PRICE CHANGE PER CENT PER YEAR, 1896–1913**

Sugar, raw	0.2%
Coke	0.5
Steel rails	0.9
Pig iron	1.0
Copper, ingot	1.0
Steel billets	1.1
Flour, wheat	1.8
Leather	1.9
Bituminous coal	2.1
Anthracite coal	2.3
Rubber	2.3
Wheat	2.6
Cotton yarns	2.6
Beef	2.6
Print cloths	2.8
Cattle	2.1
Petroleum	3.5
Cotton	3.9
Potatoes	4.3
Interest rate on call loans	—0.2
Yield on 15 railroad bonds	—0.1
Discount on 60–90 day commercial paper	0.7
Index of industrial stock prices (Dow-Jones)	3.4

Source: F. C. Mills, *The Behavior of Prices.* National Bureau of Economic Research, 1927, p. 67.

the other, both costs and prices being held in check with varying effectiveness by price controls. The disparity of wartime price movements can be illustrated by the fact that prewar (1939) wholesale prices were exceeded in 1943 by 111 per cent for drugs and pharmaceuticals, while electricity showed an average drop of 24 per cent and gas a drop of 9 per cent. Wholesale prices increased on the average of 33.7 per cent during the period.

Changes in prices over a period of time influence the relative economic positions not only of individuals but also of political and social groups. If price changes are large and widespread they affect the course of national policy and may even threaten the stability of the government. The depressed level of prices of farm products in the United States during the 1920's, followed by a further price recession in almost all lines during the early 1930's, was one of the influences responsible for the success of the

Democratic party in 1932. The hyperinflation of prices in Germany after World War I virtually wiped out that nation's middle class, because savings, insurance policies, salaries, and pensions—all paying fixed amounts in terms of money—became worthless when prices of goods rose to astronomical figures. Even in less dramatic form, changes in the relative prices of particular commodities or services drastically affect large classes of people. The decline in interest rates during the 1930's, for example, dealt a heavy blow to people whose income was derived from securities, raised the cost of private retirement annuities, threatened the finances of many endowed institutions, and reduced the interest charges on the national debt.

Without the aid of reliable measures of price change, however, it is impossible to determine the magnitude and rate of price changes, even for items of everyday consumption. It is often argued by people who remember the "good old days" that things were cheaper then; the contrasting prices of beer, haircuts, or of cigars "then" and "now" are quoted to prove the point. A similar argument, if more significant and with better evidence, is used when it is said in wage disputes that since the "cost of living" has gone up while wages have not, a wage increase is overdue. Whatever the purpose, however, it is impossible to determine the extent or rate of price change (and such attendant measures as the "cost of living") without the use of some simple statistical measures applied to accurate and reasonably complete records of prices.

To measure the change in the price of a particular commodity or service over a period of time, and perhaps in a given retail outlet, it is a relatively simple matter to record the prices of the commodity and observe directly those variations which are deemed significant. Thus, someone might record and analyze the monthly changes in the price of Lucky Strike cigarettes, in packages of twenty, sold in the United Cigar Store located at 42nd Street and Fifth Avenue in New York City during the period 1939-44. Such a series of prices would of course be of limited interest, for most people are usually interested in broader questions involving price changes. What has happened to the cost of living? Have wages gone up (or down) more or less than the cost of living has gone up (or down)? In other words, attention is usually centered on the movements of a *group* of prices which vitally affect a considerable number of people.

INDEX NUMBERS

To measure such changes one may use an *index number*, which is a numerical measure of the changes in a group of related variables (e.g.,

prices) in relation to their magnitude at some particular time. For example, the cost of living in the year 1939 may be represented by the number 100, and the index number for the cost of living in 1944 relative to 1939 may be (according to one index) 126, or 26 per cent above the 1939 level. Thus the change from 1939 to 1944 is measured by expressing the cost of living in 1944 as a *percentage* of the cost of living in 1939.

A "cost-of-living index" is used here as an example of the index number method because great attention and sometimes controversy has centered on it. During World War II, wage rate increases were limited in part by the "Little Steel formula." In May, 1942, a wage increase of 15 per cent was granted workers in the smaller steel companies. This set a pattern for wage increases throughout most industry, it being held that wage advances should be limited to 15 per cent above the level of January 1, 1941, since the cost of living had risen by about 15 per cent from that date to the time of the "Little Steel" increase in May, 1942. The index of cost of living used in these calculations was prepared by the Bureau of Labor Statistics of the United States Department of Labor. Subsequently, in January of 1944, the labor representatives of the War Labor Board issued a report purporting to show that by December of 1943 the cost of living had increased 43 per cent above the level of January, 1941, as opposed to the figure of 23 per cent given by the official Bureau of Labor Statistics index. The labor representatives thus sought to justify further wage increases. In doing so they raised a public argument over the official index which was reflected in angry "letters to the editor" and perturbation among statisticians. While the controversy aroused widespread public interest, it also revealed a considerable amount of ignorance as to the meaning of index numbers. The formation, use, and significance of index numbers, therefore, need thorough examination.

What is the purpose of any index number; what is to be measured with it? Once this question is settled, the choice of method and the limits of the use of the index are almost self-determined. Suppose the purpose is to measure changes over a period of time in the cost of food. But whose cost of food—a millionaire's, an unskilled worker's, a professional man's? Suppose it is decided to measure the cost of food of the unskilled worker. The next question is: where—in a certain town, in all cities over a given population, in the United States? Suppose the measurement is limited to a certain town. Next, it is necessary to consider whether to measure what the worker actually spends, or what he would have to spend if he bought the same things over a given period of time. This is one of the most confusing problems encountered in using index numbers such as those intended to

measure changes in the "cost of living." If what the worker actually spends is measured, several things are measured at once—changes in the prices of the same items, changes in the *types* of items bought, changes in total expenditure because of changes in income, and perhaps several other factors. Since people are usually interested in how much prices have changed for the same types of food which were bought at a previous period, this question can be answered by selecting the food bought by unskilled workers in some period, such as the year 1939, and tracing their particular prices during succeeding periods.

Tracing the prices of food presents a problem of research methods. Shall all foods be priced or only a sample of them? The Gallup, Hooper, and other polls have made the public aware of the process of sampling. If the outcome of a national election can be predicted within a margin of error of 2 or 3 per cent by questioning a fraction of 1 per cent of the population, it would seem that the sampling method is valid and can be used for price indexing with a suitable degree of accuracy. It would, obviously, be excessively burdensome to price every one of the hundreds of food items purchased by families of unskilled workers. A few dozen items, carefully selected as representative, will give a reasonably accurate measure of price changes of the total number.

When the representative sample of food items has been selected, still another problem arises: How can several dozen prices be reduced to one figure which represents all of them and which can be compared with a similar figure representing such prices at another date? The problem can be exemplified by choosing four food items in this hypothetical food price index: (1) one-pound loaf of bread, (2) five-pound beef roast, (3) one dozen eggs, (4) one pound of potatoes. Suppose now that what is wanted is a measure of how much the cost of food for an unskilled worker's family has changed between January 1 and July 1 of this year. The retail prices on the four items are shown in Table 48.

Table 48 **RELATIVE PRICE CHANGES, FOUR COMMODITIES**

	January 1	*July 1*	*Per cent change Jan. 1–July 1*
1. Bread (1-lb. loaf, white)	$ .10	$ .10	0.
2. Beef roast (5-lb., chuck)	1.50	1.25	—16.666
3. Eggs (Grade A, medium)	.50	.48	—4.000
4. Potatoes (white, 1 lb.)	.05	.07	+40.000
Total	**2.15**	**1.90**	**+19.333**
Average (Total ÷ 4)	**.5375**	**.475**	**+4.83**

The total price for the four items dropped from \$2.15 to \$1.90; and the average price (which is the total divided by four) dropped from \$0.5375 to \$0.475. Both total and average prices show a decline of over 11.6 per cent. Another measure of the change is to find the percentage change in each price, then average the percentage changes. By this method there is an average *increase* of 4.8 per cent in the prices.

Other methods might be used to obtain still different amounts of changes in the cost of food. Obviously, all methods cannot be correct, so it is necessary to select that one which is appropriate to the purpose of the inquiry. What is being measured? Again, the answer is the change in the cost of food, as represented by certain items, for an unskilled worker's family. This being the purpose, it is advisable to determine how much unskilled workers spend on different types of food.

Investigation of the budgets of a group of such families by the Bureau of Labor Statistics shows that in a recent year they distributed their food expenditures approximately as in Table 49.

Table 49 DISTRIBUTION OF FOOD EXPENDITURES

Food types	*Percentage of total food expenditure*
1. Cereals and bakery products	16
2. Meats, fish, and poultry	28
3. Dairy products and eggs	25
4. Other foods	31

(In the hypothetical sampling, the four items under consideration are assumed for ease of calculation to represent accurately all food items bought; thus it is assumed that the price movements for the four items represent the movements for the groups in which they fall.)

To measure changes in the cost of the four items to a worker's family, a change in potato prices should be given almost double the weight given to a change in bread prices, since group 4 is twice as large in the family consumption as group 1. But in the previous calculation, when *a simple average of the actual prices* on January 1 was found, bread had double the importance of potatoes, beef roast 30 times, and eggs 10 times the importance of potatoes. Again, when the percentage changes were averaged, *equal* importance was given to each item, since each percentage (relative) was added and the total divided by their number. This method, termed *a simple average of price relatives*, like the former one, is also inadequate for the present purpose.

What is really wanted, then, is an average which will accurately reflect the importance of each group of foods actually purchased in some typical period. Since the importance of each group is known, a price change in each group (as measured by the item) can be given the weight or proportion which that group has in the typical family budget. One way this can be done is to put each relative into the average not once, but enough times to give it a proper weight. Thus, the percentage change for bread is entered 16 times as opposed to 31 times for potatoes, etc.

If the January 1 prices are represented by 100, then the calculation in Table 50 can be made.

Table 50 WEIGHTED INDEX NUMBER

	Price relative for July		*Weight*		*Price X Wt.*
Bread	100.00	×	16	=	1,600.00
Beef roast	83.33	×	28	=	2,333.24
Eggs	96.00	×	25	=	2,400.00
Potatoes	140.00	×	31	=	4,340.00
Total	**419.33**		**100**		**10,673.24**
Weighted average					106.73
Simple average					104.83

By multiplying each price relative for July by its appropriate weight, a weighted total is obtained which, when divided by the sum of the weights (= 100), provides a weighted average. *The weighted average of price relatives* thus gives the answer: an increase of 6.73 per cent in the cost of a normal food budget from January 1 to July 1. This answer measures (remembering that a drastically reduced number of prices has been used for ease of calculation) the change in the *cost* of food for a typical unskilled worker's family between the two dates. The method, in effect, finds the cost of a certain market basketful of food in the two months.

Another method which yields the same result is the *aggregative method*,[1] which consists simply of finding out how much a fixed amount of some product costs in a given base period as compared with any other period. By comparing the total price paid for a fixed amount of a type of food each succeeding week with the total price paid in a given base week, an index could be constructed which measures the changes in cost of a given combination or market basket of food items. Thus, if a given basket of goods cost $10

[1] An aggregative index is used to measure the changes in prices in primary markets. See pp. 482-84.

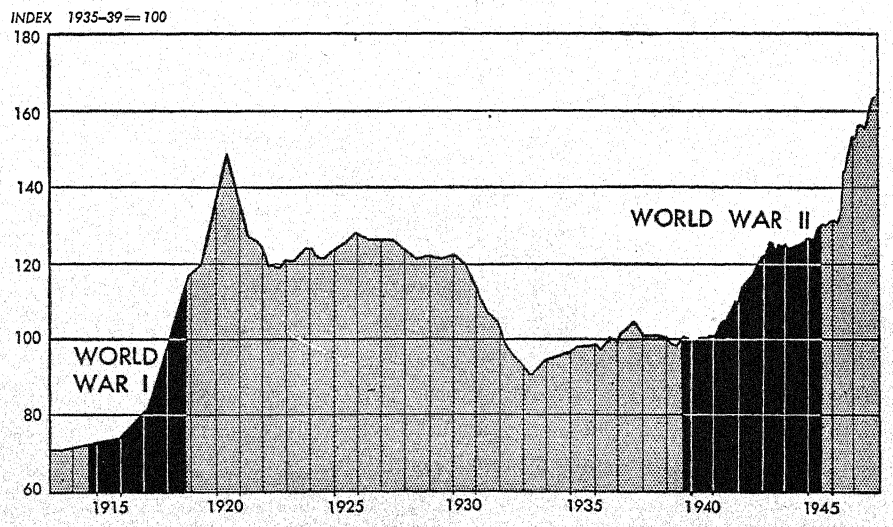
one week and $10.67 in another, it can readily be seen that the cost has gone up 6.7 per cent.

SOME INDEXES OF PRICES CURRENTLY IN USE

Consumer Price Index (Cost of Living)

The method of constructing a price index for foods which has been presented here is essentially that used by the United States Bureau of Labor Statistics in its cost-of-living or *consumer price index.* This index is designed to measure changes in prices paid for goods and services by city people with moderate incomes. Because of widespread misinterpretation of the term *cost of living,* the index is now designated as a "consumer" price index. It is based upon the prices of 183 items of commodities and services bought by urban families who in the period 1934-36 had an income of about $30 per week. Prices are collected in each of 34 large cities (56 cities are used for food), and an index is calculated for each city on the basis of the weights appropriate for that city. The city indexes are then averaged, using population figures as weights, to devise a weighted average cost-of-living index for the country as a whole. The index does not attempt to measure the changes in living costs of any

Chart 19 **CONSUMER PRICE INDEX. ALL ITEMS—AVERAGE FOR LARGE CITIES**



but moderate income groups; nor does the index measure with any great degree of accuracy the changes in the cost of living for any one group of workers, such as railroad firemen or janitors. It is a broad average.

The cost of living in the United States as measured by the Bureau of Labor Statistics Index is shown graphically on Chart 19. It should be noted that the base period (= 100) is the period 1935-39, since most people can remember conditions which existed in a recent period.

It will be noted in the chart that the two world wars have been responsible for the major upward movements in the cost of living. The peak was reached after World War I in 1920, when the average figure for the year was 143.3 in terms of the 1935-39 base of 100. Price control during World War II kept the index down to a moderate rise in terms of the base period. However, with the end of price control in 1946, the index rose sharply and stood at 164 in late 1947. A large and sustained reduction in the cost of living began in 1920, with the collapse of the war inflation, and again in 1930 after the beginning of the great depression. Major short-period changes in the cost of living have usually reflected such economic disturbances as wars, with their attendant inflation, and marked periods of prosperity and depression.

Wholesale Prices

The measure most generally used to indicate "general" price movements in the United States is the official *wholesale price index* of the Bureau of Labor Statistics. This index, which dates back to 1890 (and can conceivably be extended back to 1720 on the basis of not altogether satisfactory data) is a generally reliable measure of prices of fairly standard commodities at the mill, factory, or primary market level. By and large, the index measures the price at the level of the first commercial sale at principal points of distribution, rather than at retail outlets as in the cost-of-living index. The construction and use of the wholesale price index illustrate many of the problems of measuring price changes.

This index is at the present writing based upon price quotations for 889 separate commodities. Weights derived from the Censuses of Manufacturers and other official statistics are used to ensure that the index reflects the importance of each of the commodity groups represented by the individual quotations. The index is calculated by using the aggregative method. To obtain the individual price series making up the index an aggregate which is the product of a base price—1926 price—is multiplied by

the quantity entering the market in a recent period, and the resultant figure is compared with an aggregate which is the product of the current price of the same item multiplied by the same quantity. The method may be illustrated by three kinds of coal combined into an aggregative index in Table 51.

Table 51 AGGREGATIVE INDEX NUMBER

Bituminous coal (on tracks, destination)	*1926 price (ton)*	*Quantity weight (thousand)*	*Price X weight*	*1943 price (ton)*	*Quantity weight (thousand)*	*Price X weight*
Mine run	$4.314	227,354	$980,805	$5.045	227,354	$1,147,001
Prepared sizes	4.794	132,699	636,159	5.310	132,699	704,632
Screenings	3.422	92,844	313,712	4.207	92,844	390,595
Total			**$1,930,676**			**$2,242,228**

Coal index: 1926 = 100.0
1943 = 116.1

The index process consists of finding the *values,* at 1926 and 1943 prices, of all the bituminous coal entering the market at the average of the 1929 and 1931 *volumes.* (The average of the years 1929 and 1931 is the weight base used.) The calculations given here indicate that an amount of coal which would have cost 1,931 million dollars in 1926 actually cost 2,242 million dollars in 1943. The price index (1926 = 100) was 116.1 in 1943; in other words, coal prices had risen 16.1 per cent between 1926 and 1943. This process is repeated for all of the 889 commodities; the sum of the current prices, multiplied by the weights for all of the commodities, is expressed as a percentage of the sum of the 1926 prices, also multiplied by the weights, to obtain the index number for the current period.[2]

The course of wholesale prices in the United States over a period of almost two centuries is presented in Chart 20. As the chart indicates, the major peaks of prices have been occasioned by wars. A careful examination of the broad movements of prices over the two-century period also shows several clearly recognizable long-term *trends,* which are movements in one direction sustained over a considerable period of years. One such

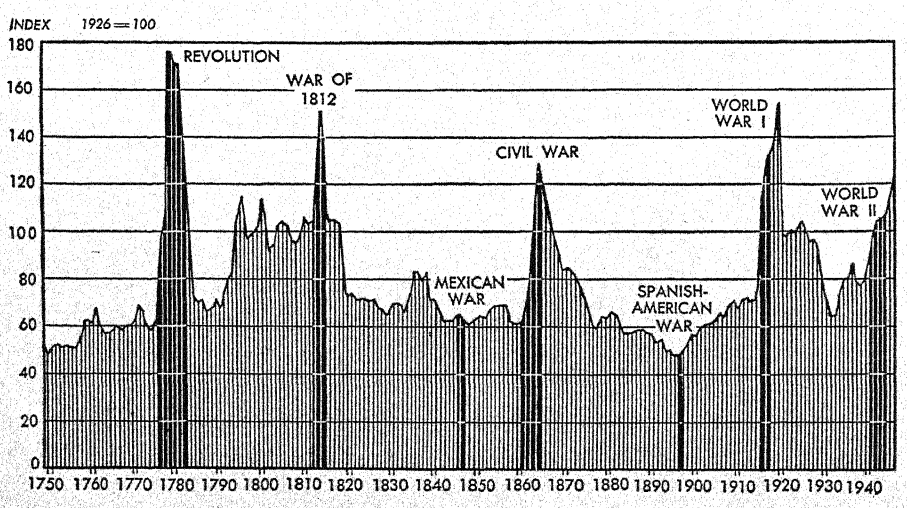
[2] If p_0 = base-year prices (1926)
p_1 = current prices
q_0 = base-year quantities (1929-31)
P_1 = index number, current period
Then the formula for *BLS* Index may be expressed:

$$P_1 = 100 \times \frac{\Sigma(p_1 q_0)}{\Sigma(p_0 q_0)}$$

trend continued upward until the period of the War of 1812, followed by a long downward trend continuing almost until the time of the War between the States, and after which another long downward trend was experienced until about 1897. From 1897 until the wartime price peak in 1920, the trend was generally upward. It again turned downward after 1920, and continued so, with violent interruptions, until the beginning of World War II.

The reasons for these long-term movements in prices are closely linked with variations in the supply of money.[3] Their economic consequences in terms of employment and unemployment, profits and losses to business concerns, and changes in the position of debtors and creditors are easily recognized. When the trend of prices is upward, business profits are easier to make, land values appreciate, long-term debts are easier to pay, and long-term creditors suffer a partial loss of the real capital represented by the loans. The economic environment is, furthermore, generally favorable to a steady expansion of employment. When the trend of prices is downward, values of all kinds of real property tend to depreciate and business profits are likely to be lower, long-term debtors find it increasingly difficult to meet interest charges and to repay loans, and creditors often acquire property by foreclosure of mortgages or by the acquisition of an ownership interest in place of the creditor interest in business corporations.

Chart 20 WHOLESALE PRICES. ALL COMMODITIES—YEARLY AVERAGE



[3] See Chap. 11, pp. 249-56.

The *trends* of prices should not be confused with the behavior of prices during business *cycles*. A business cycle, unlike the trend, is generally recognized by observers as it occurs. Trends can only be discerned from statistical records after they have been under way for a long time. Cyclical peaks and troughs in prices are scattered throughout the history of prices. The prewar peak in 1937 and the prewar trough in 1939 are clearly discernible, as is the trough in 1933 following the sharp break in prices in 1929. Other cyclical peaks in prices occurred in 1925, 1923, 1920, 1913, 1910, 1907, 1900, 1895 and at numerous earlier points in time.

Still another type of price movement, more regular and predictable than either trend or cyclical movements, is the regularly recurring *seasonal*

Table 52 **SEASONAL MOVEMENT OF EGG PRICES**

Consecutive weeks	*Grade A eggs sold to independent retailers*
1	49.4¢
2	46.4
3	45.4
4	44.4
5	43.4
6	42.4
7	41.4
8	40.4
9–22	39.4
23	41.4
24–25	41.8
26	42.8
27	43.8
28	44.8
29	45.1
30	47.1
31	48.1
32	49.1
33	50.1
34	51.1
35	52.1
36–41	53.1
42	54.1
43–50	55.1
51	53.8
52	51.8

movements in the prices of many commodities. Some of these, like the traditional January "white sales," are to a large extent customary. Most, however, reflect seasonal changes in supply and demand which occur regularly each year regardless of the level of prices at the time. An interesting illustration of the regularity of such seasonal movements was the Office of Price Administration (OPA) ceiling prices for eggs, which were calculated from the study of seasonal movements of egg prices under free market conditions. A portion of the price tables used in the regulation is shown in Table 52.[4] The maximum (delivered) prices beginning in the first Thursday in January and each consecutive week thereafter are shown here. The data are based on designated counties in certain midwestern states. Prices for sales not only to retailers but to all buyers were calculated to reflect the large amounts produced during the early part of the year and the smaller amounts produced during the latter part of the year.

Wage Indexes

One of the price measures of widest interest is the index of wages, since changes in wages are closely associated with changes in the well-being of a large part of the population. The wide interest in what happens to wages, however is accompanied by considerable confusion and misunderstanding of wage statistics. Some people are led to believe that because "wages" dropped less than many other prices during two recent depressions, labor did not suffer so severely as farmers, whose prices fell a great deal. Obviously, wages in this context must mean "wage rates," and probably mean "union wage rates"; and the failure of such rates to change has little to do with how much income workers lost because of unemployment, shorter work weeks, down-grading, shifts to lower paying jobs, and similar causes. The various meanings of wages must, therefore, be considered before analysis can be based on wage indexes.

When wages mean the price of labor services (as the term was used in the foregoing chapter on wages), the term applies to hourly or piece rates of pay and similar types of wage payment. A wage as a price is the amount paid for doing a particular job or working at something for a given period of time. However, wages also constitute incomes, and in the sense of incomes, wages must not be considered as rates but rather as earnings per week, per month, or per year. The use of wages as a measure of well-being is something else. The well-being of workers depends not only upon wages in monetary terms, but also upon what the money wages will buy. Well-

[4] Revised *Maximum Price Regulation* 33, "Eggs and Egg Products."

being depends further upon nonwage factors which cannot be reduced to monetary terms. Shorter hours, regularity and pleasantness of work, freedom to use leisure time, and community services such as free education, public health service, and recreation facilities—all add to labor's well-being.

Because wages is a term having several meanings, it is not surprising to find different meanings of the word employed in recent postwar wage disputes. Labor leaders are wont to call attention to the rise in wage *rates* as smaller than the rise in the cost of living (or the cost of a certain standard of living) and also by way of invidious comparison, as they appear in relation to the rise in industrial profits. Others have compared the rise in labor's weekly or annual *earnings*—which reflected overtime rates and steady employment—with the rise in cost of living and the rise in commodity prices. During depressed periods of business, labor leaders tend to avoid comparisons of wage rates with the cost of living, since the index of the latter usually falls faster than rates; instead, they stress the decline in weekly, monthly, or annual earnings resulting from shorter hours, unemployment, and similar causes. Their opponents may then stress the fall in the cost of living relative to wage rates, and especially the drop in prices and profits relative to the drop in wage rates.

Because the number of job classifications is enormous and comparison between them difficult, data on wage rates are more difficult to collect than data on earnings. As a result, the most readily available measure of changes in wage rates over any extended period of time is average hourly earnings, which are usually derived from reports on total pay rolls and employment (more recently, pay rolls and man-hours worked). Changes in hourly earnings do not accurately reflect changes in wage rates because (1) changes in overtime worked may cause spurious changes in the average; (2) up-grading or down-grading causes a change in the average without a change in rates; (3) changes in the proportions of lower and higher paid workers within the labor force affect the average even though rates may not have changed. Over a long period, however, average hourly earnings data indicate fairly well what has happened to the average price of labor, or wage rates.

The movement of average hourly rates from 1820 to 1940 is shown in Table 53. Average money rates of wages in American industry more than doubled between 1820 and 1890, and more than quadrupled between 1890 and 1940. There is no doubt that money rates have been trending strongly upward for more than a century, although there were setbacks after both the War between the States and World War I as a reaction to high wartime levels. There were setbacks also during several major depressions, particularly during the depression beginning in 1929.

Table 53 AVERAGE MONEY AND REAL HOURLY EARNINGS IN U. S. INDUSTRIES, 1820–1940

	Average money wages		Average real wages	
	1913 = 100	1914 = 100	1913 = 100	1914 = 100
1820	36		41	
1830	37		51	
1840	41		51	
1850	43		59	
1860	47		57	
1870	84		71	
1880	66		77	
1890	74	69	96	93
1900		75		99
1905		83		101
1910		91		98
1915		100		102
1920		231		111
1925		225		131
1930		236		142
1935		240		175
1940		296		212

Source: Based upon studies by Alvin H. Hansen, 1820–1890; Paul H. Douglas, 1890–1926; and the Bureau of Labor Statistics after 1926.

One of the reasons for the phenomenal eightfold rise in money earnings from 1820 to 1940 is to be found in the enormously increased productivity of labor during this period. The opening up of new lands, the tapping of newly discovered resources, and the use of greatly increased proportions of capital in relation to man power, coupled with revolutionary technological advances, served to increase the marginal productivity of labor and, therefore, to raise wage rates. However, not all of the rise in money wage rates can be attributed to greater productivity, when productivity is taken to mean *physical productivity*. Between 1896 and 1920, for example, the general trend of prices was upward and the value of product in terms of money was increasing. Much of the advance in money wage rates during this period was made possible by the rise in prices. Between 1920 and 1930, on the other hand, money wages rose very little, but since prices fell the maintenance of money wages represented a very considerable increase in labor productivity. Likewise, the more rapid rise in wages than the rise in prices from 1933 to 1940 could not have been possible if the higher wages had not been supported by labor's productivity.

If changes in money wage rates are taken as a measure of labor's well-being, the picture is somewhat different. The purchasing power of average hourly money earnings may be considered as one measure, although a poor one, of such well-being. An increase or decrease in money wages often does not indicate an increase in the purchasing power of such wages, because if money wages rise less than the cost of living increases, then average rates are worth less than before. Thus, if a wage of $1 buys $1 worth of goods at one time, and the wage is raised to $1.50 while the goods increase in price to $1.25, then $1.50 will buy only one and one-fifth times as much as $1 bought before.

The purchasing power of money wages in terms of cost of living commodities is termed *real wages.* An index of real wages can be derived by dividing a money wage index by a cost of living index. An index of real wages is presented alongside the index of money wages in Table 53. Between 1820 and 1890, it will be noted, real wages increased generally in proportion to the rise in money wages. However, between 1870 and 1880, when money wages dropped, real wages increased because the cost of living had dropped more than money wages. Although money wages rose between 1905 and 1910, real wages dropped slightly. Most important, however, is the difference between the growth in money wages and the increase in real wages between 1890 and 1940. Money wages quadrupled, but real wages at the later date had risen only to somewhat more than twice the 1890 level. Such great divergences indicate that data must be used carefully in the study of economics. A measure of wages must be selected which is appropriate for the purpose of the study.

INTEREST RATES

Numerous measures of movements in the prices of loans, i.e., interest rates, are quoted in almost every daily newspaper. Call-money rates, rates on commercial paper, rates on Treasury bills and certificates, and similar measures of short-term rates—all are readily available but nevertheless tend to reflect more or less ephemeral activities. At various times short-term interest rates have reacted sharply to such happenings as gold imports and exports, the movement of crops to markets, tax collections, issuance of securities by the government, or the collapse of stock-market values. Short-term rates over a period of years are likely to show marked fluctuations which are reflected only slightly in the prices people must pay for the money loans which are destined to finance the acquisition of capital goods. A somewhat more satisfactory measure of fundamental movements in interest

rates is the yield on long-term bonds of high quality. Data for almost a century on long-term bonds are given in Table 54.

Table 54 YIELD OF HIGH GRADE BONDS, 1860–1942*

	American railroad bonds	*High-grade corporate bonds*
1860	6.25%	
1865	5.40	
1870	6.65	
1875	5.62	
1880	4.64	
1885	4.11	
1890	3.58	
1895	3.52	
1900	3.20	3.30%
1905	3.45	3.50
1910	3.73	3.80
1915	4.17	4.15
1920	4.91	5.10
1925	4.48	4.50
1930	4.39	4.40
1931	4.08	4.10
1932	4.73	4.70
1933	4.17	4.15
1934	4.08	3.99
1935	3.52	3.50
1936	3.36	3.20
1937	3.15	3.08
1938		3.00
1939		2.75
1940		2.70
1941		2.65
1942		2.65

* Railroad bond yields are from Table 10 of F. R. Macaulay's *Some Movements of Interest Rates, Bond Yields, and Stock Prices in the United States since 1856*, National Bureau of Economic Research, New York, 1938. The January figure for each year is quoted. Corporate bond yields are the 30-year basic yield from David Durand's "Basic Yields of Corporate Bonds, 1900–1942," Technical Paper 3 of the National Bureau of Economic Research, 1942. Yield figure is for the first quarter of each year.

Although bond yields are influenced by many factors—capital imports and exports, the state of bank reserves, and the level of national output, among others—there is considerable correlation between bond yields and wholesale

prices when the long-term rate of interest is subject to free market forces. This is illustrated by the downward trend in both bond yields and wholesale prices from the close of the War between the States to the middle 1890's, and the rise in both from 1900 until about 1920. One hypothesis advanced to explain this movement is that money yields automatically reflect real yields. As wholesale prices fall, bond prices tend to rise and their yields consequently tend to fall; as prices rise, bond prices tend to fall and their yields to rise. In terms of price analysis, the supply schedule for long-term loans may tend to increase when prices are falling and to decrease when prices generally are rising.

The fairly consistent downward trend in interest rates since 1930, however, reflects more or less conscious management of the money market, especially since 1933, which has been explained in relation to government finance.[5]

PARITY PRICES

Probably one of the most significant manipulated changes in the national price structure has been the legislation embodying the concept of parity prices for agriculture. The sharp decline in farm prices after World War I focused attention on the disparity between the prices of things farmers sold in relation to the prices of what they bought. An effort was made to secure "equitable" (upward) adjustments in farm prices to preserve or restore the economic prosperity and well-being of the farm population. During most of the period between the two world wars (in wartime farm prices achieved a high degree of inflation), the idea of *parity prices* was refined, reduced to statistics, preached, and accepted into the economic folklore of the volatile-priced agricultural economy. In 1933, after considerable promotion, it was embodied in legislation (Agricultural Adjustment Act) designed to raise farm prices. Parity prices have been a standard fixture in national policy on agriculture since that time. During World War II, the standard doctrine of parity prices did not permit sufficient inflation of farm prices to suit the friends of the farmer, however, so for price control purposes several higher variants were legislated to prevent parity prices from restraining further inflationary increases in farm prices.

The basic principle of parity prices is that "fair" prices for farmers require the maintenance of a certain relationship between the prices received by farmers and the prices paid by farmers. If the prices paid rise more or fall less than the prices received, it is assumed that agriculture's position in the

[5] See Chap. 19, pp. 442-43.

economy has deteriorated. Since the parity concept was developed in 1922 after a severe drop in farm prices, it was natural that the pre-World War I days when farm prices were higher should be considered a fair base. Thus, parity for most crops has been defined as a purchasing power for farm products in the period August 1909-July 1914. Special exceptions have been made for certain crops—for tobacco the base period was shifted to 1919-29, and then to 1934-39, as these periods became successively more favorable to tobacco growers.

The general index for parity can be calculated from indexes of prices paid by farmers and prices received by farmers. If, on the basis of 1909-14 = 100, the index of prices paid at a particular time stands at 150 and that of prices received at 100, farm prices are obviously below parity. To put it simply, $1 worth of purchased goods in the base period would now cost $1.50, whereas the farm products with which to buy them would still sell for $1. Farmers would, under these conditions, get only 66⅔ per cent of parity.

For a particular crop, for example wheat with a base period (1909-14) average price of $0.98, parity can be calculated by multiplying the current index of prices paid—say, 150—by the base period price. The parity price of wheat would be $1.47 ($0.98 × 150). If the actual price were less than $1.47, it would be less than parity. The formula is simple, and its appeal to agriculture is attended not only by words of approbation but by the most important test of a politico-economic policy—votes. The way this formula has been used to determine government subsidies to agriculture will be discussed in the next chapter.

The parity concept, despite its settled place in national policy, is full of economic mischief both in what it is and in what it may become. Fundamentally, it uses relative prices as a measure of economic well-being, whereas a more appropriate measure for this purpose is real income. Parity gives no recognition to the improvements in farm productivity by means of which income per unit of farm labor or capital has greatly increased. Farm prices may well fall below parity without reducing real farm income, because improvements in farm productivity may reduce costs more than prices. Parity takes no account also of the increase in social income shared by farmers through the development of good roads, free farm services, subsidies to agricultural education, and the like. As an instrument of national policy, by concentrating attention and government support upon farm prices, the seeking after parity has had serious repercussions upon export markets for American cotton and other export crops, which have been partially driven out of foreign markets by cheaper foreign sources of supply. This unfortunate consequence of domestic price-support activity

has led to the inevitable next stage—dumping and subsidization of exports. It might require ultimately the subsidization of domestic manufactured products made from the high-priced raw material if domestic manufacturers are to remain competitive in world markets.

For what it may conceivably become, the concept of parity is also disturbing. Organized labor, now represented by talent comparable to that championing agriculture and possessed of as great facility for getting out the votes, has half seriously suggested that parity wages for labor might become a worth-while goal. Select, for example, some base period in which wages were particularly favorable in terms of the cost of living, and seek to restore wages to that favorable relationship. It follows almost by definition that a favorable base period for labor will be one during which farm prices are low, since farm prices exercise an important effect upon the cost of living. There would be, then, a parity wage for labor inconsistent with parity prices for farmers. To the extent that farm prices rise toward parity, wages will tend to fall below wage parity. But since both groups want increased incomes, the pressure upon prices in general will be upward. The possible effect might resemble a dog chasing his tail up a stairway.

Finally, of course, there might develop a movement for parity prices for industry, whose base period would be disadvantageous to both labor and agriculture and whose parity prices would be inconsistent with both farm and labor parities.

Perhaps in recognition of these actual and potential drawbacks of farm parity, responsible officials of the Departments of Agriculture and State suggested as an alternative early in 1945 that farmers might be paid a subsidy directly rather than indirectly by means of price manipulation. That agriculture has suffered more severe price reductions than most of the other parts of the economy during periods of business depression cannot be doubted, as the evidence on price inflexibility shows.[6] If, as a matter of national policy, the government is determined to protect farmers from the rigors of periodic and uncontrolled price deflation, it appears that a choice must be made between supporting prices and supporting farm income. Price support, the policy pursued in the period between the wars, raises grave problems of equity, leads to dumping of surpluses of farm products, impairs export trade, and tends to be self-defeating for export crops. Moreover, it can be argued that the maintenance of high prices of certain farm products encourages uneconomical use of resources, i.e., land may be used to grow crops which might be more economically used for something else—grazing

[6] Described in the following discussion.

or timber growing, for example. Income subsidies rather than price support might reduce uneconomical use of land.

FLEXIBLE AND INFLEXIBLE PRICES

A type of price relationship which has caused almost as much controversy—both political and economic—as the parity price relationship for agriculture is the relation between "flexible" and "inflexible" prices. This relationship, which had been noted even by Adam Smith in his *Wealth of Nations,* was "rediscovered" and put forward as a cause for the great depression which began in 1929. It was also part of the politico-economic diagnosis which gained some favor during the 1930's, based on a belief that market influences no longer proved an efficient means of regulating the economic system.

The essence of the flexible-inflexible price controversy is illustrated in Table 55 showing the relative declines in prices and in production, between the prosperous year 1929 and the depression year 1933, for selected industries.

Table 55 **PERCENTAGE DECLINE IN PRICES AND PRODUCTION IN SELECTED INDUSTRIES, 1929–33**

	Per cent drop in prices	*Per cent drop in production*
Motor vehicles	12	74
Agricultural implements	14	84
Cement	16	55
Iron and steel	16	76
Automobile tires	25	42
Leather products	33	18
Petroleum products	36	17
Textile products	39	28
Food products	39	10
Agricultural products	54	1

Source: National Resources Committee, *The Structure of the American Economy,* Pt. I, Table XXII, Washington, 1939.

It was contended that the American economy was divided into two types of industries: (1) those that maintained prices at the expense of production; and (2) those that were unable to maintain prices and, therefore, maintained production. The industries at the top of Table 55 showed the smallest reduction in prices and the largest drop in production; those at the bottom

showed large reductions in prices and little drop in production. The inference (based on an inaccurate application of demand and supply analysis) was that some industries cut their prices, or were forced to do so, and thereby maintained production and employment; whereas other industries chose deliberately to hold their prices high, and therefore cut their production to the extent necessary to maintain prices. An attempt was made to show that the high price-low production behavior was characteristic of those industries where production was concentrated in a few big producers while the low price-high output behavior occurred where production was divided among many producers. The villain in the piece, therefore, became "industrial concentration," the control over prices and production achieved by a few corporate giants. The concentrated industries behaved in such a manner that competitive industries like agriculture suffered from low prices of their products and high prices for their trucks, tools, equipment, and supplies, while labor and the country as a whole suffered from unemployment and a low level of income.

A careful analysis of these allegations will serve to illustrate some of the pitfalls involved in applying economic principles without fully understanding them. Consider first the implication that the drop in production was caused by maintaining prices at too high a level. Prices, as it has been reiterated, are the result of forces operating on both the demand and supply sides of the market. An important factor influencing the price at which goods will be sold is their cost of production. A study of changes in direct costs of production for 86 industries for the period 1929-33 showed that the direct cost of production dropped most for those industries which showed the greatest drop in price, and least for those that showed the smallest drop in price.[7] Obviously, motor-vehicle prices could not drop in price as much, for example, as food products, when the direct costs—largely labor and materials—of making motor vehicles fell but little, whereas the direct costs of food—based upon prices of agricultural products—dropped enormously. The differential reduction in direct costs of manufacture, of course, requires an explanation itself. An explanation is found in the change in prices of productive services.

In agriculture and to some extent in extractive industries, prices for capital, for the services of land, and even for much of the labor are almost all like rent. The price of land, of course, is rent, but once capital and a family career have been committed to agriculture, the return received at least over a short period of years depends on the same forces which deter-

[7] A. C. Neal, *Industrial Concentration and Price Inflexibility*, American Council of Public Affairs, Washington, 1942.

mine land rents—i.e., what the market will support. Moreover, since there is little alternative use either for farm land, farm capital, or even farm labor other than for working farms, crops continue to be made regardless of price. In many cases, this tendency to maintain output is accentuated by the necessity to meet mortgage payments and to pay taxes and insurance, all due in terms of stated sums of money. When prices of farm products fall, one common way of obtaining a given sum of money for such payments is to increase production enough to offset the reduction in price. Corporations in manufacturing might miss dividends and even go into receivership without much change in physical production or even in employment, but farmers who go bankrupt lose not only their businesses, but their homes, life work, and self-respect as well.

In addition to the enormous declines in prices for agricultural products and such commodities as timber products which are produced under the influence of similar economic forces, movements in the prices of other productive services help to explain price inflexibility. Industrial wages in many lines are relatively inflexible, particularly in strongly unionized industries. Average hourly earnings in 44 major manufacturing industries, for example, dropped 13 per cent from 1929 to 1932. Public utility rates either rose or changed very little during the depression period.

A final test may be applied to the theory that "concentration" caused prices to be maintained at the expense of production. If this is the case, the extent of concentration should be related on the one hand to the flexibility of prices, and on the other to the drop in production. An analysis of the drop in prices from 1929 to 1933 in relation to the extent of concentration for 407 commodities showed no perceptible relationship. That is, a large decline in price was equally likely to be associated with high or low concentration. Similar results were obtained when the degree of concentration was compared with the depression drop in production.[8]

There is abundant evidence that the explanation of price inflexibility in terms of a sinister choice either to maintain prices at the expense of production or to maintain production at the expense of prices, is not supported by good economic reasoning or investigation even though it has had its political uses.

Far more significant than the investigation of "inflexible" prices from the standpoint of maintaining income and employment is the influence of "concentration" upon business savings. There is some evidence that those

[8] W. L. Thorp and W. F. Crowder, "Concentration and Product Characteristics as Factors in Price-Quality Behavior," *American Economic Review, Papers and Proceedings*, V. 30, 1941. Also "The Structure of Industry," TNEC Monograph No. 27.

industries which were relatively more concentrated succeeded far better than the less concentrated industries in maintaining the margin between direct cost and price which provides the source of business savings. Index numbers of the margin between average direct cost and average price for 67 industries divided into low, medium, and high concentration groups show the pattern given in Table 56.[9]

Table 56 **INDEX NUMBERS OF COST-PRICE MARGINS**

	1929	1931	1933
Index for low concentration group (22 industries)	100	76.8	68.3
Index for medium concentration group (22 industries)	100	84.4	75.1
Index for high concentration group (23 industries)	100	92.0	85.7

The size of the margin between direct cost and price dropped twice as much in the more competitive industries as in the highly concentrated industries. Although these figures show only the drop relative to a prosperous year (1929), they do indicate that concentration exercised some influence on unit margins over direct cost during a period of depression. It is likely, on the basis of these data, that "concentration" exercised its effect, if any, upon national output and employment by maintaining the basis of business savings during a period when investment outlets were scarce. If the volume of savings arising in one part of the economy is excessive in relation to investment opportunities, other parts of the economy must dis-save; this dis-saving may have to be effected through a general reduction in national income.

If producers behaved in a purely competitive manner, prices would fall to the level of average direct cost when demand fell off sharply. As a result, the unit margin over direct cost would tend to disappear. Since the margin over direct cost provides the source of business savings and—through salaries, interest and dividends—the source of a large part of private savings, it is clear that purely competitive behavior by business would tend to correct any tendency for savings to become excessive. On the contrary, if savings were sufficiently reduced by reductions in the margin over direct cost from price-cutting, a relatively small amount of new investment would tend to restore the level of the national income.

Price movements, therefore, are central to the formation of the level of

[9] Neal, *op. cit.*, p. 137.

national output and national income. Changes in the structure of prices, meaning changes in some prices relative to others, are indicative of changing demands, changing techniques of production, and changing patterns of income distribution. They are a means of bringing about a redeployment of resources among uses and a reallocation of income between savings and spending consistent with the maintenance of a high level of national output.

CHAPTER 23

Price Control

Conscious and purposive control over prices is undertaken because the conditions established by market forces are regarded as undesirable, either to sellers or to buyers. Usually price control is instituted for single commodities or groups of related commodities, although at times it is thought desirable to control virtually all prices. Control may be exercised by buyers, by sellers, or by public authority in the form of national, state, or local governments.

CONTROL OF INDIVIDUAL PRICES

Private Control of Prices

There is perhaps much truth in Adam Smith's canny observation, made over a century and a half ago, that "people of the same trade seldom meet together, either for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices." Control over prices is an ever-present ambition of most sellers under the system of free private enterprise; private sellers are seldom satisfied that the prices they can get are, to their way of thinking, "fair," especially if they are prices established in a competitive market. As a result private sellers employ many methods to achieve some degree of control over their own prices.

An individual seller may attempt to achieve monopoly power for himself

by several devices. He may obtain ownership or control over virtually the whole capacity to produce a particular product, as was done before 1941 in the United States in the production of aluminum, nickel, molybdenum, magnesium, shoe machinery, glass-container machinery, scientific precision glass, telephone service, and Pullman cars.[1] Or again, if a single firm is unsuccessful in controlling prices, two or three firms in duopolistic or oligopolistic industries may each have some degree of control over prices. Duopoly was approximated in the prewar period in the sale of domestic telegraph service, bananas (at the import level), plate glass, electric-light bulbs, electric accounting machines, railroad air brakes, compressed oxygen and acetylene, oxyacetylene torches, and sulphur. Oligopoly prevailed in a large number of industries, including steel, cement, agricultural implements, newsprint, glass containers, typewriters, eyeglasses, radio broadcasting, automobiles, electrical equipment, rayon, and many others. Uniform pricing practices, such as the use of basing points or zone-delivered prices, are also methods giving some control over prices.

Attempts by private interests to control prices are not limited to businessmen. Farmers have for many years sought to achieve some control of crop prices by means of co-operative marketing agencies. One of the more successful of these has been the California Fruit Growers' Exchange, which controls the marketing of most of the California citrus fruit crop. The exchange withholds fruit from the market by channeling it to processors or dumping it when the supply is so great that uncontrolled selling would unduly depress prices. Similar co-operative marketing schemes have been organized for cranberries, California prunes and apricots, California lettuce, fluid milk sold in several metropolitan markets, and numerous other farm products. This kind of organization was legally sanctioned by the passage of the Clayton Act of 1914 and encouraged by subsequent legislation.

The collective-bargaining practices of labor are, essentially, monopolistic means to affect the prices of labor services. The most important objective of unionization is control over the supply of labor of particular kinds or in particular industries. Collective bargaining has been sanctioned and encouraged by legislation for many reasons, mainly because employers in dealing with unorganized labor have a considerable degree of monopsony power that may be used to beat down wages below competitive rates. Unionization tends to counterbalance the employer's power, but it may also lead to monopolistic practices which exploit employers, at least for a time.

[1] See Clair Wilcox, "Competition and Monopoly in American Industry," TNEC Monograph No. 21.

PUBLIC CONTROL OF PRICES

Until the early 1930's, public policy toward control of individual prices was based upon a generally accepted conception that industries were either "competitive" or "natural" monopolies. Control of prices in the competitive area was deemed to be maintained by the forces of the market, although beginning in the 1890's the public was led to demand enactment and enforcement of antitrust laws designed to keep competition free and effective. For monopoly areas, however, it was early recognized that control of prices could not be left safely to market forces; the Supreme Court in the Granger Cases of the 1870's officially recognized that regulation of the rates of "natural" monopolies might be necessary. Railroads, street railways, water supply, electric power, gas, telephone, and telegraph systems—all have by now been judged to be natural monopolies which must be subjected to regulation if they are to serve the public most efficiently. The public control of such "natural" monopolies is discussed in Chapter 30.

Since the early 1930's the clear-cut classification of industries into two groups—"natural" monopolies and competitive industries—no longer afforded a guide for public policy toward price control. The coal-mining industry, one of the most competitive, has been judged to require minimum price control. (Unlike most controls upon public utilities and railroads, control of coal prices was directed primarily to prevent price cutting.) Prices of many agricultural commodities have also been subject to some form of minimum price control. In fact the outstanding example of selective price control outside the traditional sphere of natural monopolies is that exercised over agricultural commodities.

Control of Farm Prices

Government intervention to raise farm prices following 1929 was, as already noted, the immediate result of political pressure from farm areas. Ultimately, however, it was caused by an accumulated series of maladjustments dating back to World War I. Farm prices soared from 1913 to 1919 under the impact of wartime demand; by 1919 farm prices were on the average more than double their prewar level. Stimulated by high prices and government urging to increase output, many farmers bought additional land at high prices by mortgaging their property, so that some new land was brought under cultivation. In 1920 farm prices collapsed with the onset of a general depression beginning in that year and remained somewhat depressed

throughout the otherwise prosperous 1920's. Meanwhile, European agriculture was rehabilitated and several foreign countries undertook to protect their own markets by restricting imports. In addition, American dietary habits became more diversified, to the detriment of markets for traditional staple products like wheat, corn, and hogs. Rayon cut into the cotton market, and the growing use of gasoline-driven automobiles and tractors reduced the demand for hay and feed grains.

In general, American agriculture after World War I was characterized by an enormously expanded capacity to produce, a large debt burden contracted when prices were high, and a reduced demand for its products. This economic condition was aggravated and in a sense continued by the fact that the demand for many crops is inelastic; smaller crops will generally produce more revenue than larger ones. The supply of most staple crops is likewise fairly unresponsive to reduced prices. Faced with the necessity of meeting payments upon loans, taxes, and similar expenses involving fixed annual outlays, many of the nation's farmers respond by increasing output when prices fall in order to compensate for lower prices by increased marketing.

These underlying economic forces led to constant agitation to "do something for the farmer," which meant to do something to raise farm prices. The traditional method of affording some degree of price protection to American producers is the tariff, which had been a part of America's economic policy since the days of Alexander Hamilton.[2] A tariff on farm products, however, provides little benefit to farmers because in general the United States is an exporter, not an importer, of farm products. A general increase in tariff duties in 1920 and 1922 and again in 1930 probably hurt American agriculture more than it helped, because it reduced foreign purchasing power for American goods (by reducing imports into this country) and led foreign countries to raise further barriers against American exports of farm products.

Direct price support for farm products began with the passage by Congress of the Agricultural Marketing Act in 1929. The Federal Farm Board was established under the act to advance loans to farm co-operative associations, which in turn advanced loans to farmers on the security of their crops. The loans were designed to enable the farm co-operatives to hold crops off the market until prices were more favorable. Unfortunately, operations were started in 1930 at the beginning of another landslide in farm prices, and the Farm Board was soon required to set up stabilization corporations which rescued co-operatives by taking over their holdings of farm products at the loan rates. Despite these actions, together with open-market purchases of wheat and cotton by the government to bolster prices, the

[2] See Chap. 36, pp. 788-95.

world depression carried crop prices to new low levels. Eventually, the Farm Board had to liquidate its holdings at a loss of about 340 million dollars. This experience demonstrated that in the face of a world-wide depression it is impossible to support prices determined by the world market without control over production.

The Agricultural Adjustment Act of 1933 created the Agricultural Adjustment Administration (AAA) and set a goal for price-supporting operations: to re-establish parity between farm and nonfarm prices. To achieve this goal the government made cash ("benefit") payments in return for agreement by the farmers to reduce their acreages [3] of certain basic crops such as wheat, cotton, corn, hogs, tobacco, and rice. In order to achieve immediate results, the now-famous slaughter of little pigs and the plowing-under of one fourth the cotton crop was conducted in 1933, its purpose being to reduce amounts produced and so force increases in prices. For other farm crops, the AAA assisted the farmers in drawing up "marketing agreements" with the purchasers of the crops, such as canneries. The intent behind the marketing agreements program was also to raise the prices received by farmers.

Although the legislation was declared unconstitutional in 1936 (in a case involving the tax levied on processors to raise funds for benefit payments to farmers), Congress in the same year enacted the Soil Conservation and Domestic Allotment Act as a constitutional means of achieving the same purposes. Instead of receiving payments for not producing the staple crops, farmers under the new act received payments for producing something else. They were paid for producing soil-building crops like legumes instead of wheat, corn, cotton, and other soil-depleting crops.

The fall of farm prices in the second half of 1937 led to the passage in 1938 of an Agricultural Adjustment Act authorizing the "ever-normal granary." The object of the "ever-normal granary" was to carry the excessive crops of some years into other years of poor crops and good prices. In order to avoid the experience of the earlier Federal Farm Board, however, the "ever-normal granary" plan continues the soil-conservation payment plan set up under the 1936 act, and furthermore provides for national voluntary acreage allotments for the basic crops—wheat, corn, cotton, tobacco, and rice—and authorizes loans to farmers for carrying over crops from good years to bad (low- to high-price years). Excessive amounts of the basic crops can be held off the market by imposing marketing quotas if two thirds of the

[3] Since the farmers were required to reduce the number of acres planted, the reduction in the size of the crop was not necessarily proportionate, because the tendency was to put the poorest acres out of production.

farmers participating in a referendum approve. When quotas are imposed, sales in excess of quotas are subject to penalties. Parity payments in excess of soil conservation payments may also be made whenever the price of one of the five commodities falls below parity, assuming of course that Congress appropriates the funds necessary to make such payments.

These several price-raising measures have been accompanied by other measures to aid the farmer which are to some extent contradictory to them; cheap credit is made available to farmers overburdened by mortgages, needy farm families are aided in attaining a self-supporting basis on the land, tenant farmers and farm laborers are encouraged with loans to buy farms of their own, and sound methods of farm management are urged on farmers to make farming more profitable. Although these additional programs have undoubtedly benefited the farm families for which they were developed, it is probable that they have also served in some measure to maintain and increase farm output. There can be little doubt that the price-raising measures—even before World War II—succeeded in increasing farm income, or furthermore that the gains enjoyed by farmers were to a considerable extent offset by the higher prices paid by consumers and the heavier obligations placed upon taxpayers to meet the cost of the programs to the government.

During World War II the demand for agricultural products increased greatly because of the higher level of national income, the increased individual consumption of food by persons in the armed forces and the large shipments abroad under Lend-Lease. This increase in demand raised farm prices to about twice their prewar low levels, and protection was afforded farmers by wartime legislation which forbade imposition of ceilings at less than 110 per cent of parity. In order to encourage farmers to increase food production without fear of depressing prices, even if the war suddenly came to an end, legislation guaranteed "support" prices of 90 per cent of parity for corn, wheat, rice, tobacco, and peanuts for two years after the end of the war (the termination date of December 31, 1948, is established on these guarantees). Other crops added later to the "postwar price-floor" program by Congress included hogs, eggs, most kinds of chickens, turkeys, milk, butterfat, certain varieties of dried peas and beans, and potatoes. Cotton was guaranteed 95 per cent of parity. To take care of any surpluses that might arise to threaten price support policy, Congress provided that they might be sold abroad at "competitive world prices."

The attempts of government to raise farm prices since 1929 offer a good example of the manner in which initial attempts to regulate prices lead to additional and continued regulation. The Federal Farm Board's attempt to

raise prices was foredoomed to failure because it neglected control of supply. Subsequent legislation broadened the control to include production quotas, implemented by appropriate penalties for enforcing such quotas. Since the domestic supply of the major crops constitutes only a part of the world supply, domestic price-raising programs reacted against foreign sales. In the 10-year period after 1927-28, the American share of the foreign cotton market dropped from almost 50 per cent to only 25 per cent. It therefore became necessary to separate the world price structure from the domestic, the latter being maintained at higher levels and protected by tariffs and prohibitions against importation of previously exported American farm products. The "price-support" program late in 1945 had reached a state in which domestic prices to producers for the major crops were guaranteed regardless of market influences. The provision authorizing foreign sales at world "competitive prices" raises other problems. Since farm products are exported not only in raw form but are also embodied in manufactured articles, American exporters were confronted with the unpleasant prospect of buying their raw materials at higher domestic prices than their foreign competitors. To remove this disadvantage, the government faced the unpleasant task of rebating to thousands of manufacturers the difference between the domestic and world prices for the agricultural raw materials used in that part of their production which is exported.

Even such an expansive program, however, cannot undo all the damage to the export position of manufactured commodities. Higher prices for major crops will undoubtedly affect prices for minor crops also, since the two types of crops are in competition for the land, labor, and capital necessary to grow them. Acreage controls for major crops may increase output of minor crops and lead to control on them too. Thus, in a myriad of hidden and unsuspected ways, maintenance of domestic prices for farm products above world levels will work mischief in foreign and domestic markets. For these reasons and because dumping of surpluses abroad is inconsistent with the announced American policy of removing restrictions upon international trade, it is likely that other means of raising domestic prices without destroying the export market will be sought. One such means is an extension of the use of international commodity agreements.

International Commodity Agreements

International commodity agreements are agreements among producers (or their governments) in different countries to control the production or marketing of particular commodities. The main purpose is to support world

prices for commodities important in international trade. A few examples of peacetime commodity agreements may be cited to illustrate their operation.

British rubber producers joined in 1922 in a scheme to restrict output in order to maintain prices (the Stevenson Plan). The plan, to the extent that it succeeded in raising prices, encouraged an expansion of production by Dutch and native plantations not bound by the agreement. It came to its logical end in 1928 after British producers had been forced to make successive reductions in their own output to offset unregulated Dutch production. A restriction plan was tried again in 1934, under government auspices and including all major producers. Production quotas were established, based upon past exports, to maintain an agreed-upon price. The scheme was continued with considerable success until Japan occupied the principal producing areas in 1942, and it was again put into operation after the defeat of Japan. A similar plan was tried by private producers of tin in 1929, but failed because some major producers refused to join. In 1931, this time under government auspices, restriction was agreed to by all major producers and was actively practiced. Output was limited by production quotas and export quotas were set under the plan in accordance with past exports. In effect the restriction plan led to the liquidation of excessive stocks and an improvement in prices. Brazilian producers of coffee, who account for the major part of the world's supply, established a price stabilization scheme in 1923, which involved the storing of coffee during years of bumper crops in order to guard against depressed prices. Production control was not exercised and a succession of bumper crops led to large stocks. The scheme broke down, and the government was impelled to intervene with a program of destruction of surplus crops coupled with subsidization of the growers. Coffee was again controlled in 1941 when the Inter-American Coffee Agreement became effective. Under this agreement export quotas are set for each American producing country for the purpose of allocating exports between the United States, the principal consumer, and the rest of the world. The United States, whose interest is solely that of a consumer, is represented on the board administering the agreement. Sugar, tea, and wheat are among the other commodities for which agreements have been established. In general, the trend in all such agreements is toward increased control by each government over its domestic production, stocks on hand, and uses of the commodities; for experience has shown that control over price is ineffective without regulation of production.

In general it will be found that most international commodity agreements originate with private producers, not governments, and are intended to stop or reverse a downward movement in prices. Governmental intervention is

invited partly to lend its authority to make production control effective. Such agreements may be mutually beneficial to producer and consumer alike. Prices for many agricultural and mineral raw materials entering into the world market have been subject to extreme fluctuations. Rubber, for example, has ranged in price within a few years from 4 cents to about $1 per pound. If international control can be exercised to provide a greater degree of stability for such prices, and with such stability removal of uncertainty from both producers and consumers, it might serve useful purposes. Stability of raw-material prices would do much to stabilize incomes in small nations heavily dependent upon exports of raw materials. Notwithstanding this, however, the prices thus stabilized should be high enough only to bring forth the output demanded. Under such a type of control there would appear to be little need for the elaborate provisions limiting production to specified quotas except in those cases where a marked and permanent shift in demand required a reduction in productive facilities. Under ordinary circumstances, buffer stocks would be bought up and held off the market in years of high output and released during years of low output while prices would remain relatively stable. Such a plan is almost foredoomed to failure, however, because prediction of output and demand is exceedingly difficult and, more important, because it is likely to be politically impossible to stabilize prices at low enough levels to make production control unnecessary.

Control of Bituminous Coal Prices

Control of minimum prices was undertaken during the depression of the 1930's in the bituminous coal industry for some of the same reasons that price support was provided for agriculture. During World War I new mines were opened under pressure of the greatly increased war demand and the resulting high prices, and output was considerably increased. During the years between wars demand shifted markedly toward other sources of fuel. Before World War I coal was the source of about 70 per cent of the total energy used in this country; by 1937, it was the source of only 45 per cent, owing to the increased competition of fuel oil, natural gas, and water power. Moreover, substantial economies in the use of coal were effected. Together, these changes restricted the demand for coal and the high-cost producers suffered losses.

Wages in the bituminous coal industry are a large proportion of total cost, usually more than half. Price-cutting in the industry resulted in wage-cutting which undermined the welfare of almost a half million coal miners. By 1933 the annual earnings per miner had been reduced to the meager sum

of $550. Since coal mines are located in more or less isolated areas, alternative opportunities are not readily available for unemployed or underpaid miners; furthermore, most miners are not trained for other work. Low wages did not remove the oversupply of mine labor, nor did low prices close down enough mines to remedy the chronic condition of overcapacity which developed during the 1920's. Long-run adjustments, it should be observed, take a long time to occur.

Under these conditions, coal operators and miners united in pressing for legislation to provide relief from low prices. Legislation was passed permitting the establishment of minimum prices for coal, and, for a time, the fixing of wages and hours by government. The task of fixing prices in the coal industry was by no means an easy one. A brief description of the process is given here to illustrate the complexity of government price control. All coals produced by all mines had to be listed in terms of a uniform system of classification so that the prices established could be related to particular specifications. Since coal prices are quoted on a delivered-to-market basis, f.o.b. mine prices had to be made consistent with the delivered price system. Seasonal price patterns had to be recognized and provided for, since coal prices are normally reduced in the summer to encourage sales. The competition of competing fuels had to be recognized and, if possible, the loss of markets avoided. After taking these and other considerations into account, the minimum prices were required to return to each of the 10 major producing areas the weighted average cost of production for the mines within it.

Like much other legislation aimed at supporting prices, the Bituminous Coal Acts attacked the symptoms rather than the malady of low prices. Minimum price control legislation made no provision for closing down mines unable to cover average total costs but on the contrary helped directly to keep them operating; it did not serve to transfer men from coal mining to other occupations but instead encouraged miners with higher wages to remain in the occupation. Although the legislation did make conditions more attractive for both operators and miners by increasing their earnings, this was accomplished by raising the price of coal to the public. In fact, by specifying average cost as a standard for minimum prices the legislation provided the basis for possible future trouble, since prices by this standard would probably be increased if demand and output fell. Thus the government had, wittingly or unwittingly, sponsored a device which even the more ruthless monopolists have hesitated to use. With World War II, this legislation was allowed to lapse.

Control of Fluid-Milk Prices

Fluid-milk prices were under minimum price control in 20 states by 1940. The devices used to control minimum prices for milk during the 1930's provided a basis for maximum price control during the war, and in the post-war period minimum price controls have once again been established in many areas to encourage regular production and to insure orderly marketing.

Before the depression of the 1930's milk markets could be characterized as monopsonistic on one side and monopolistic on the other. The principal buyers from the producers were large distributors, whose principal market was composed of unorganized family consumers and retail stores. Likewise, producers in most important milk marketing areas were organized into co-operatives for marketing their milk and had thus improved their bargaining position with distributors. The depression carried fluid-milk prices down along with other farm prices, however, and the influence of marketing co-operatives in protecting producers' prices was largely eliminated by the underselling of nonmembers.

The fluid-milk price structure existing before the depression (and later adopted by price-control authorities) consisted generally of a scheme of class prices for the same type of milk, different prices being paid for the same milk depending upon its end use. The most simple system is a two-price one: Class I milk—all milk and cream sold for human consumption in fluid form; and class II milk—all milk not sold as class I, including milk and cream used for making ice cream, butter, cheese, condensed milk, and other products. More elaborate systems of classifications are found but need not be discussed here.

Control of the minimum prices of milk, like control of other farm products, was undertaken especially after 1933 to maintain or to increase the prices paid to producers. The milk producer occupied a position in the structure of his industry similar to that of the coal miner in the bituminous coal industry—pressures upon prices at the level of final sale were transmitted through the levels of distribution eventually to bear on him and thus reduce his earnings. Federal government control is effected by entering into marketing agreements with milk distributors, by issuing licenses, or by establishing minimum prices for milk entering into interstate commerce. Orders may be issued only if approved by producers accounting for two thirds of the total number or two thirds of the volume of production. This latter provision gives to producers' co-operatives effective veto power over such orders, since co-operatives control the major proportion of the sales.

Since federal control is limited to milk shipped across state lines, the scope of federal authority is greater in the Northeast, where milk markets are large and states small, than in the West and South. Most states include many purely local markets not subject to federal control; consequently federal control in practice is maintained in co-operation with those 20 states exercising milk price control.

Under both state and federal regulation, control over prices is exercised in well-defined "market areas" made up usually of urban territory. The minimum prices specified in some areas are based upon costs, but in others the primary basis appears to be the market prices previously existing plus a considerable amount of bargaining among producers, distributors, and government agencies. In most markets, the state milk control boards have established minimum resale prices for fluid milk to protect the structure of minimum producer prices. In this process they usually reduced or eliminated the spread between store and home-delivery prices, despite the obvious differences in cost of distribution by the two methods. As might be expected, the use of the power of the Federal and state governments to enforce minimum prices was instrumental in raising producer prices and therefore received the support of producers, distributors, and their employees.

It is generally agreed that some control of milk production and distribution is in the public interest because the production and distribution of milk vitally affects the health of most citizens. Clean, pure, and wholesome milk is almost as essential to the public as pure water, which has long been the accepted concern of government. Establishment of minimum prices for milk is often defended as essential to the maintenance of an adequate supply of pure milk because low prices are believed to encourage the sacrifice of quality for quantity. Regardless of these contentions, several considerations must enter into the establishment of minimum prices. An abnormal spread between the prices of class I milk and other classes must result in the subsidization of manufactured milk products by fluid-milk consumers. The class price system in its more refined form appears to be little more than a device for monopolistic price discrimination, the effect of which is to maximize returns by charging what the traffic will bear. The lowest price market must at least return out-of-pocket cost of production or it is being subsidized by other markets. Establishment of minimum resale prices for milk encourages wasteful competition in distribution. Since distributors cannot cut prices to increase their sales, they are led almost inevitably to use other methods—excessive advertising, maintenance of high-cost routes, too frequent deliveries, and other costly services. When store sales and other methods of distribution are discouraged by minimum price-fixing,

the consumer is given no choice as to whether he wants more service or lower prices; thus the cost of distribution tends to become frozen at high levels.

Control of the Prices of Productive Services

It should not be judged from the extensive consideration of commodity price control that government price control is limited only to commodities. It has already been shown that the level and structure of interest rates is dominated by the Federal government's monetary and banking policies. Direct control over interest rates is for these reasons seldom necessary. However, maximum rates are prescribed in about 30 states for consumer loans where the amount borrowed is under $300. Rates are enforced to curb the avarice of the small-loan business which, before legislation, often took advantage of the financially distressed borrower requiring small sums for emergencies.

Another example of government control is found in the wages of workers involved in interstate commerce, which have been subject to minimum standards under the Fair Labor Standards Act (1938). State laws devoted to the same purpose were initiated in 1912 by Massachusetts. Minimum wages were specified for many industries in the short-lived codes of fair competition administered by the National Recovery Administration from 1933 to 1935. Control over wage increases during World War II was exercised by the War Labor Board.[4]

DIRECT CONTROL OF THE PRICE LEVEL

The reasons for direct control by the government of *individual prices* stand in sharp contrast to the reasons for control of the *price level.* Individual prices or groups of prices are controlled by government to prevent gouging of consumers in the case of monopoly price control, and in almost all other cases prices are controlled by government to aid the producers affected. On the other hand, control of the *price level* by direct methods is possible only by government, for it alone has the extensive authority to undertake such control.

Control over the price level is typically undertaken during a wartime period, although it may be attempted during some other period when the economic system is in danger of breaking down. Thus, in both world wars direct control over the price level was instituted, and at the depth of the

[4] These forms of price control are discussed in Chap. 29.

Great Depression serious attempts were also made to bring the level of prices under control. The more dramatic and complete type of direct price control used during World War II to prevent inflation will be discussed first, since the partial controls used in an attempt to stop deflation in 1933 represented to a considerable extent the same forces working in the opposite direction.

Wartime Price Control

The fundamental reasons in both world wars for governmental attempts to control the price level were (1) a breakdown of the price system in some areas, and (2) the unwillingness of government to allow it to operate in the remaining areas. During war the demand of government for goods of all sorts, when superimposed upon civilian demands, results in a rapid expansion of employment. A very rapid rise in prices begins as soon as full employment is approached. Such a rise in the price level, however, serves no useful purpose, because under full employment conditions higher prices all around will not induce additional production.

The government, moreover, does not allow the price system to work in its usual fashion even in those remaining areas where it still may perform its economic function of allocating resources and apportioning consumer commodities. High prices on some items will encourage their production to the exclusion of other items, and only those persons able to pay the inflated prices can buy the goods. But this may be directly contrary to the interests of the nation's war effort. For example, high prices for butter may encourage diversion of dairy products, feeds, cattle, and farm resources to this use, to the detriment of urgent needs for other milk and animal products. Moreover, high prices for butter and other food items may lead to discontent among workers and ultimately to disruptive and delaying strikes for higher wages. It follows that when extensive controls are placed over wide areas of an economy it soon becomes necessary to make them universal, for the normal operation of the price system in any significant degree tends to invalidate the objectives of the control.

During war a disproportionately large share of all productive activity is devoted to the output of goods for military purposes. Wage payments and other income payments initiated by the production of war goods constitute purchasing power not matched by a corresponding value of civilian goods. When a large part of war goods production is financed by funds borrowed by the government from the commercial banks, the disparity between spendable income and the amount of civilian goods becomes large and inflation

threatens. On the other hand, if government financed the war in such a way —by exceedingly high taxation—that consumer expenditures were no larger than the value (at prewar prices) of the current output of consumer goods, direct control of the price level would be unnecessary. Only certain items in abnormally short supply would need to be subjected to price control, and even these might be left free of price control if they were carefully rationed. Such a system of war finance cannot be attained because it would require a level of taxation which would be politically impossible to achieve and which might also impair the incentive to produce for both workers and businessmen.

In this country and in the other belligerents, therefore, taxes and the gross savings of the economy fell short of the amounts of money required by government for war expenditure; the difference was made up by the sale of bonds to commercial banks, which pay for them by creating deposits in favor of the government. The amount of disposable income available for purchases of civilian goods is therefore greater than the amount of such goods available at prevailing prices, in consequence the level of prices must rise, and the same amount of civilian goods will necessarily cost more.

These conditions are almost always incident to wars owing to the impossibility of levying high enough taxes or inducing sufficient saving. Direct control of the level of prices is the outcome. Several features of this policy may be noted:

1. *Financing the war is made easier.* If prices are held down, not only is the amount of borrowing by the government less, but also consumers are unable to spend as much as they would like and are forced to save or buy bonds. Moreover, assurance that prices will not rise encourages people to save; fear of rising prices encourages them to spend.
2. *The financial cost of the war is lessened.* Once inflationary increases in prices begin they have a tendency to undermine careful purchasing by procurement agencies. Wages and the cost of all materials entering into war goods will increase, causing constant revisions of government contracts and the eventual result that producers will insist upon charging a risk premium to compensate for possible further price increases.
3. *The financial burden of the war is distributed more equitably.* A free price system operating under a condition of inflationary finance will lead people with larger incomes to bid up prices of goods needed by all income classes; rising prices force the poor to restrict consumption more drastically than the rich. Since dependents of people in the armed forces must live on fixed allotments, they would suffer and

military morale would ultimately be affected by inflationary prices.

4. *Direct allocation of labor and materials for war use is facilitated.* Allocation of resources by price—the traditional method—is supplanted during wartime by an allocation based upon military requirements. Imposition of price control gives free rein to direct allocation, which can be implemented when necessary by consciously directed price manipulation.
5. *Transition to the postwar period is eased.* There is no collapse of prices at the end of the war, since prices are not at inflated levels. Reconversion of industries converted to war production can take place more smoothly, since large adjustments in prices of goods taken out of production by the war are rendered unnecessary.

Control over the price level may take the form of *selective direct controls* or of *general direct control.* During World War II the United States exercised selective control from the fall of 1940 to May of 1942, it being recognized that full employment and full utilization of capacity would not be reached in all industries or all areas at the same time. Consequently, as supply became short in particular cases, voluntary agreements and price schedules were put into effect. In essence this method was designed to control prices of particular industries or products to ensure an orderly rise in the price level conducive to the attainment of full employment. Bottleneck items needed in war production were first brought under control. The first price-control schedule issued was, appropriately, on secondhand machine tools needed to retool plants for war production. Controls multiplied until 55 per cent of the wholesale value of manufactures, minerals, and imported commodities were under price control early in 1942.

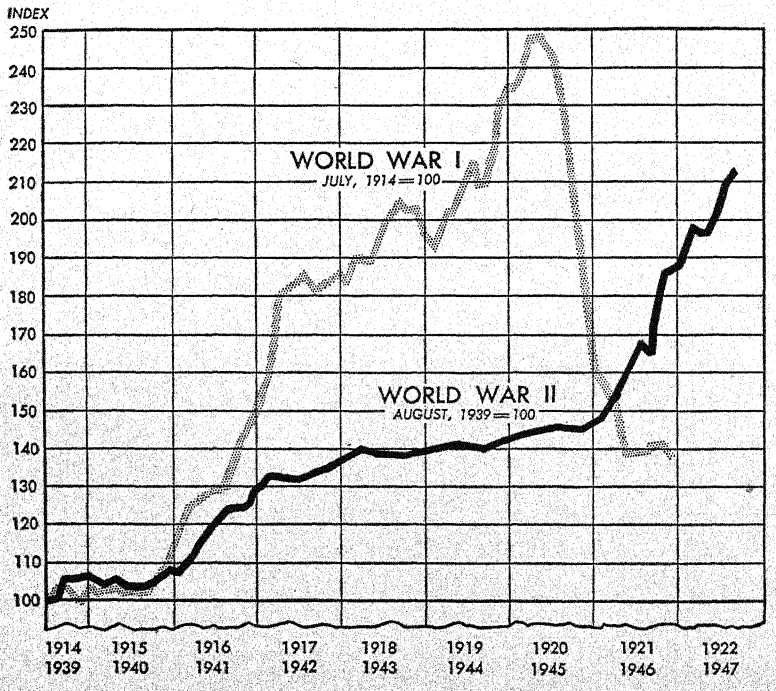
Faced with growing inflationary pressures on uncontrolled commodity prices the OPA adopted general price control with the issuance of a General Maximum Price Regulation in April, 1942. This regulation froze the prices of 23 per cent more of the wholesale value of all commodities (leaving only about 22 per cent uncontrolled) and instituted control widely at the retail level. This regulation was partially replaced later by more specific regulations covering particular industries. Commodities not price-controlled were mainly farm and food products. As already noted, Congress placed farm products in a favored category when it enacted price-control legislation. Effective control was authorized only toward the end of 1942. The effect of this delayed action is shown by fact that from May to October, 1942, uncontrolled wholesale farm products increased in price 3.4 per cent and uncontrolled food prices rose 9.4 per cent, as compared with

a rise of only 1.2 per cent in the wholesale price index for all commodities. From October, 1942, when full statutory authority to control farm prices was obtained, until the end of the war in August, 1945, the wholesale price index was held to a relatively small increase of less than 7 per cent.

The over-all record of the most comprehensive wartime price-control program in the history of the country is demonstrated in Chart 21. The chart shows the rise in wholesale prices during World Wars I and II, the base for each period being the month prior to the outbreak of the war. The inflationary peak of World War I, which was not reached until 1920, was almost 250 per cent greater than the level at the beginning of the war. The peak of World War II up to the end of 1945 was held to only slightly more than 40 per cent above the prewar level. Price controls were relaxed after August, 1945, lapsed temporarily in July, 1946, and were abandoned in November, 1946. Prices rose sharply after this date.

Although a variety of methods of controlling prices was actually used, a method in itself did not ensure maximum prices which were, to use the language of the Price Control Act, "generally fair and equitable." Tests

Chart 21 **WHOLESALE PRICES IN AND AFTER WORLD WAR I AND WORLD WAR II**



were devised to determine whether maximum prices were unduly oppressive. Since the test used by business itself in measuring its own success or failure is profit, it was expected by the government that profits would afford a measure of the adequacy of a level of maximum prices established by its regulations. The principal test used to determine whether the level of prices for an industry was fair was the relationship of current over-all profits of the industry to over-all profits in the period 1936-39. The fact that this period had been specified as a base for the purpose of calculating excess profits taxes probably influenced its selection.

The industry earnings standard, as this standard was called, was used in the following way: If an industry requested a general price increase on all or at least the most important lines of its products, current and base period over-all profits were first ascertained by the OPA. If a group of representative firms which were in business in the base period were, for example, enjoying current profits (before federal income and excess profit taxes) of 5 million dollars per year and made only 3 million dollars per year in the base period, 1936-39, a general increase was denied.[5] If, on the other hand, current profits had fallen to less than 3 million dollars, a price increase was granted with the intention of allowing at least that amount of profits to be earned.

Where the application of this industry earnings standard was inappropriate because base period earnings were not typical, or because a large proportion of the producers would suffer losses from its application, a secondary standard was often used. This standard required maximum prices to be set so as to allow profitable production by suppliers of at least 75 per cent of the output of the industry. A price increase was made whenever the producers of 25 per cent, or in some cases less, of the total output were showing over-all loss. Other standards were applied in some cases, especially in cotton textiles where the method for raising prices was virtually specified by law.

From this brief review it may be concluded that the standards most widely used for industry-wide price increases were influenced more by notions of equity than by economic principles. There was no assurance that a product or line of products would be produced merely because the industry making the product or line was making profits. For in many instances profits were derived almost entirely from the production of war

[5] Base period earnings were always adjusted for subsequent increases in investment. Thus, if investment had increased by 10 million dollars since 1936-39 and the return on investment was 5 per cent during that period, base period earnings would be increased by $500,000. This procedure in effect protected the base period return on both new and old investment.

items, many of which were not subject to price control. The industry earnings standard was, therefore, supplemented by a "product standard" which bore a close relationship to a well-known economic principle. It was previously stated [6] that ordinarily a producer will make a product if it will sell for more than his direct cost of making it. The product standard required the establishment of a maximum price high enough to cover the factory cost of at least three quarters of the total output of the industry. Factory cost was considered the accounting counterpart of direct cost.

It can be argued—and it was argued in many heated discussions between representatives of government and of industry—that the product standard represents a misplaced analogy. As an economic principle, the conclusion that it pays to produce so long as direct costs are covered is derived from a case in which the firm produces only one product. The product standard was used on producers who produced many products under conditions in which labor and materials required to make the low-priced line might be obtained only by reducing the production of higher profit lines. Under the circumstances, many producers avoided production of the low-profit (or loss) items if they could do so.

Furthermore, prices were very often adjusted not upon an industry-wide basis but by arrangements with individual sellers. These individual adjustments varied with the profitability of the seller, the minimum adjustment being factory cost. In some cases, adjustments were given to products simply because strong representations were made by the production agencies or the armed services that no other method was available to secure the output necessary. Political pressure on occasion proved too strong to be overcome, and adjustments, whether justified or unjustified, were forced. Agricultural and forest products, and products made of agricultural raw materials, were more often subject to political pressure than other products.

Since general price control by definition almost destroys price as the means of allocation, direct control of production and consumption were substituted during the war in considerable measure for the usual allocation accomplished by prices and profits. The War Food Administration established production goals for agriculture and directed the movement of foods and farm products to areas and industries experiencing shortages. The War Production Board limited the output of nonessential goods; it channeled materials to essential users and, in extreme shortages (aluminum, copper, and steel), directly determined how much material could be used by individual producers and for what purposes. The OPA rationed a long list of consumer commodities, including meat, butter, fats, oils, sugar, gasoline,

[6] See Chap. 14, pp. 320-23.

fuel oil, shoes, canned foods, stoves, and new automobiles. Direct allocation and direct control of production and consumption were, however, by no means complete, and where there was no direct control the price system was allowed to function in conformity with price ceilings.

Under wartime controls producers still had considerable latitude for the exercise of choice. Those who had war contracts often had to decide whether to produce civilian goods as well as war goods. And among civilian goods the choice was often between high-priced and low-priced goods. For example, low-priced cigars, cheaper grades of children's clothing and shoes, etc., virtually disappeared as producers concentrated on the more profitable lines. The resultant shortage of many low-priced commodities led ultimately to the "maximum average price" type of regulation (particularly for textiles and clothing) under which a producer was required to produce low- and high-priced lines to the extent necessary to maintain some calculated average price for his whole production.

Another device widely used by producers to maximize profits under price control was to stop selling to certain customers who were entitled to a low price under price control. Because the maximum price to certain large buyers was lower than the price to other buyers, the larger buyer often had difficulty obtaining supplies to meet the wartime expansion of demand. Since price control could not prevent the acquisition of sources of supply, large buyers often assured themselves of supplies by buying out their sources. Chain groceries bought packing houses to obtain meat supplies, and large packing houses in turn bought out box plants and their suppliers to obtain shipping containers. Again, in the case of upholstery fabrics, many mills had higher prices for sales to retailers than for sales to furniture factories; and, as a result, many consumers were obliged to obtain upholstery fabric for their furniture at retail. Clothing manufacturers bought textile mills, and vertical combinations mushroomed in many other lines. Assurance of a source of supply, and in some cases desire to obtain extra mark-ups allowed distributors, were the basic motives for such integration.

In the field of rent control, owners of rented dwellings found that the uncontrolled market for property was more favorable for selling than for renting their houses. The supply of dwellings for rent was thus reduced, while houses were available for purchase by people who largely would have preferred to rent; evictions and threatened evictions by new house owners forced tenants to buy houses of their own though they might prefer to rent. To curb this practice and lessen its disruptive effect it finally be-

came necessary to require six months' notice before a family could be evicted from a rented dwelling by a new purchaser.

Sellers managed in many other ways to obtain higher prices for their products, despite price control but without flouting regulations. Price is more than a dollars-and-cents figure. It almost always includes credit terms, services, return privileges, and minimum quality standards; often it includes advertising allowances, promotional material, bonuses, etc. The consumer in wartime soon found that he was not getting so much for the old price as before. Credit was restricted by government regulation, service was reduced, and generally merchandise was sold on a "take-it or leave-it" basis, so that the consumer received less although the physical quality of the good might remain unchanged. All these influences, coupled with outright violation of ceilings through black-market operations, served to raise "real" prices considerably more than official indexes indicated.

It was through hidden price increases, disappearance of low-profit lines from the market, and diversion of shipments to customers and locations whose ceilings were highest, that the fundamental disparity between supply and demand made itself felt. For almost every desirable civilian good it was found that demand exceeded supply at the ceiling price. If the seller could increase his return by dropping any services or extras or by selecting his customers, the market would support his effort. Direct control over production and distribution—including consumer rationing—was thus a necessary complement to effective price control, since these direct controls forced a reduction in the amount which could be bought at the maximum prices at one time to the level of the supply made available to local areas. Since relatively few items were rationed at the retail level, the excess of demand over supply for unrationed goods resulted in several familiar wartime conditions—queues, withholding of goods for favored customers, and intense bartering of scarce goods.

Despite the general complaints of business against controls the volume of production and profits reached during the war indicate that even with innumerable mistakes and frictions, business generally fared well under price control. Total corporate profits before taxes in 1944 were estimated to be four and one-half times the 1939 level, and profits after taxes were estimated to be more than twice the 1939 level. There is little doubt, however, that individual producers and individual industries were often squeezed by price control; it is axiomatic that if price control does not hurt anyone it is not price control. Government regulation generally—including price control—imposed a heavy administrative burden upon small businesses not able to hire specialists to handle the "red tape" of control.

Apparently the hurt that was done was not severe, however, because the volume of business failures during the war was but a fraction of its normal peacetime level. The rapid increase in prices following the termination of price controls was indicative of their effectiveness.

Direct Control to Prevent Deflation

It has been shown that an inflationary condition is characterized by an excess of the amount of income available for spending over the amount of goods available for purchase at current prices. A deflationary condition is opposite to this, being characterized by an excess of the amount of goods available for purchase at current prices over the amount of the income used for expenditure upon them. Obviously, a condition of general deflation results from some kind of short-circuiting of income, since the output of goods must of necessity create income.

A deflationary period is usually brought about by a reduction in the demand for capital goods. If the demand for money loans to be used for the purchase of capital goods falls off sharply, or if the supply of such loans is rapidly constricted, national output and income are likely to be reduced because people who were formerly employed in capital goods construction lose employment.[7] Whatever its cause, once a deflationary movement begins, the reduction in income quickly brings about a wave of price reductions in industries which sell their products in competitive markets and buy their raw materials and labor services in competitive markets. In the face of falling prices, many new capital goods projects are deferred by industry, and purchases of homes, furniture, automobiles, and other forms of consumer durable goods are deferred by many consumers. This behavior accentuates the price-cutting activity which has already started, and the general level of wholesale prices and even of retail prices begins to move downward.

By 1933, in the United States a deflationary movement had been in process for over two years. The number of unemployed had reached a total of about 13 millions; wage rates continued to be cut. Under these circumstances, the National Industry Recovery Act (NIRA) was passed in June, 1933. It was in effect a companion measure for industry and trade of the Agricultural Adjustment Act. The NIRA was directed toward the general problem of recovery and was not primarily directed toward con-

[7] See Chap. 6 for an analysis of the cumulative downward movements of output, income, consumption, and saving. The present discussion stresses the changes in price attending such declines.

trolling prices. One of its objects was to stop wage-cutting and to increase wages and buying power. Whether intended or not, one of its results was to encourage unionization of labor by guaranteeing (in Section 7a of the act) the right of labor to organize and bargain collectively. One of the purposes of the act was also to put a stop to the price-cutting which had been forcing the price level downward, since most of the unemployment and economic distress was associated by the public with the falling level of prices.

Price control was exercised under the act by means of "codes of fair competition," proposed by industry and approved by the President. Industry was exempted from the operation of the antitrust laws under the codes of fair competition. These codes were for the most part drawn up and administered by industry groups, usually trade associations. The price-fixing provisions in the codes were not directly authorized by law, but were derivative from the power to establish rules of fair competition. With respect to legal authority, administration, and enforcement, price-fixing under the National Recovery Administration (NRA) was on much weaker ground than under OPA with respect to maximum prices: the results of the two programs reflected these essential differences.

Price control under the NRA codes took the form either of direct controls over prices or indirect controls likely to influence prices. Only 38 codes contained provisions for the direct control of minimum prices by industry groups, and only in a few cases did the NRA itself promulgate minimum prices to correct demoralized markets. One of the most widely used provisions for price control in the codes was the "open price system" under which the seller was required to file a price change with a code authority or one of its subgroups. This provision was usually administered with another one requiring a waiting period to elapse before a price change could become effective. Price control under the open price system was dependent upon the amount of moral suasion exercised by competitors and, more importantly, on the effect of equalizing prices by making price changes subject to industry surveillance.

Another widely used provision was a general prohibition of sales below cost, this usually being the cost of the individual seller. Unless such a provision was accompanied by a careful definition of what constituted cost, it was likely to be ineffective. Furthermore, since costs vary among producers, this prohibition was usually accompanied by permission to high-cost sellers to reduce their prices to the level of those of their competitors; by such provision high-cost producers were protected from being frozen out of the market.

A large number of codes contained provisions regulating prices by reference to terms of the sale other than the price charged. Maximum discounts were specified, premiums limited, credit terms defined, and allowances for advertising, for trade-ins, returns, etc., limited. As in the control of maximum prices, in minimum price regulation the terms of sale must be considered as much a part of the price as the money price itself.

Since indirect control of prices can also be achieved by control of supply, this method was not neglected in the NRA codes. In some cases, the number of hours which a machine or plant might run per week or per month was specified; in others, production quotas were set and production for stock was limited. These same devices, it will be recalled, have tended to enter into international commodity agreements designed to "stabilize" prices. Finally, to prevent the ultimate defeat of the government's price-fixing program, a number of codes provided that approval must be secured before new productive capacity could be constructed or used. To the extent that this last provision was effective, it retarded investment and delayed recovery.

In both conception and administration, the NRA codes were designed to stabilize and to raise prices by means of the concerted action of private sellers operating under government supervision. It was hardly surprising, therefore, that the Board of Review, appointed by the President to appraise the NRA program, came to the conclusion that the NRA codes had fostered monopoly, had discriminated against small business, and to the extent that the codes were effective, had exploited consumers. It will long be a subject for debate whether the NRA actually did succeed in its short life in promoting recovery earlier than it would have occurred otherwise. By 1935, its program was in such disrepute and violations of codes so widespread that when the act was declared unconstitutional in the Schechter poultry case, no serious attempt was made to revive the price-fixing powers except in two or three industries characterized by chronic overcapacity. The anti-trust laws were returned to full force; "industrial self-government" was returned to its traditional position as something to be prevented rather than encouraged by law-enforcement measures.

PRICE CONTROL AND PLANNED ECONOMY

Government control of individual prices, so long as it is limited to a relatively small segment of the economy, is not inconsistent with the allocation of resources and the direction of production by the price system. The principal standard which must be met by individual price control is

whether with price control the goods and services subject to it will be produced and sold at a lower cost. Clearly, in some cases consumers would pay lower or higher prices for goods or services without price control, but the lower or higher prices might result in higher costs to the public because of exploitation of labor, unhealthful working conditions, or some other preventable burden. Thus, control of milk prices is based in part upon avoiding the costs to society of impure milk; and price support for farm products is based in part upon the desirability of having a large farm population, because it is deemed to add political stability to the country and because it tends to be the source of a considerable part of the increase in national population. "Social costs" are of course hard to measure, and the principal danger involved in the extension of individual price control by government is that it may be used primarily to protect vested interests of labor, of capital, or of landowners at the expense of efficiency and therefore at the expense of the entire population.

Government control of the price level, on the other hand, is to be regarded as an expediency to be undertaken during times of emergency, and is usually justified by the course of events before it is undertaken. It is not proposed until the need for it is obvious, and it is therefore usually imposed too late.

Government control of individual prices and of the price level as it has been practiced in the United States must be distinguished from price control which takes the form of *price-fixing*. The latter type of control sets actual selling prices, whereas the former sets only maximum or minimum prices. Maximum or minimum price control leaves to individual sellers a considerable amount of flexibility. It leads automatically to the removal of controls when they are no longer necessary. Direct price-fixing, in contrast, sets prices at which sellers must sell. It does not encourage the diminishing or removal of controls, because it is usually invoked in the operation of a planned economy designed to replace a capitalistic, competitive economy. Direct price-fixing is therefore an instrument either proposed for or used in a system of socialist economic planning.

Proponents of a planned economy sometimes argue that free choice to consumers and free choice of occupation—the economic counterparts of political democracy—are possible if the planned economy uses a price system. If it does not use a price system, consumers cannot be allowed free choice in purchasing, workers cannot be allowed free selection of jobs, and producers cannot determine whether their products are preferred to other products which might be made with the same use of resources. It has been shown that the comparison of selling price with cost, and additions to out-

put with additions to the cost of productive services, are necessary to determine what shall be produced by a firm, how much shall be produced, and what combination of factors shall be used in production.

Sophisticated proponents of socialist planning have therefore agreed that a price system is necessary under socialism. They have agreed further, in a rare tribute to the efficiency of a properly functioning competitive society, that in setting prices they must follow the competitive standard. Prices must be set at such levels that supply will equal demand both for goods and for productive services. If supply exceeds demand at the fixed price, the fixed price must be lowered, and if demand exceeds supply, the price must be raised. Individual plants under socialism must produce up to the point where marginal cost is equal to price, and must combine their factors of production in such fashion as to achieve the least cost combination. These prescriptions in a socialist plan will be recognized as the same results of pricing as are obtained under purely competitive conditions.

However, socialist planning would achieve the competitive result by radically different means than those used in a freely competitive private enterprise economy, and the difference in means vitiates the pretensions of free choice.[8] In a planned economy price changes would be made *consciously by a central controlling group,* whereas under a competitive system of private enterprise they are made almost automatically and impersonally through the operation of market forces. Private sellers in a competitive system need know very little about supply or demand or why either has changed; all they need to know is that prices have fallen or risen, adapting themselves accordingly by their own independent choice. In a centrally planned economy the controlling group must direct a change in price, and the effectiveness of its decision is dependent on the amount of knowledge about supply and demand of many products it possesses, as well as on the conformity of opinion on what ends are desirable. It is doubtful whether any group could achieve the degree of omniscience required to perform the work of keeping countless millions of prices in proper relationship to each other. Moreover, if the problem of obtaining the necessary knowledge were actually solved, there still remains the problem of making the necessary price changes in the face of political pressures which can be assumed to exist in a socialist society even as in a capitalist one. People who accept the impartial dictates of the changing market are likely to rise up in wrath against bureaucrats who so much as propose to impose a price cut upon them in order to reduce supply. Apart from these problems of operating a planned economy the point is that the freedom of choice associated with

[8] See Chap. 42 for a further discussion of this point.

a free market economy cannot exist under a planned economy, for prices themselves must be planned by central authority.

The method to be used in socialist planning for directing production also differs from that used in a competitive private-enterprise economy. Although both systems use prices, the socialist system must measure quantities prior to changing prices. Under socialist planning, if prices are too high for a particular good, inventories will pile up; if a wage rate is too high, there will be more applicants for jobs than positions open; thus the planners must keep careful records of all inventories and all applications for jobs. When inventories of goods or applications for jobs of particular kinds are abnormally high, prices or wages must be cut—by some indeterminate amount. The task—and cost—of maintaining such records of physical inventory as would be required would be formidable. Under a competitive private-enterprise system the price adjustment is made automatically as inventories begin to pile up or applicants for jobs begin to multiply. It is possible that the time required to obtain action in a planned economy might result in such shortages or surpluses of particular types of goods or labor that larger price changes would be required simply to offset the tardiness of action.

SUMMARY AND CONCLUSION

Price control limited to individual goods or services is undertaken by private interests in order to obtain advantages which an unregulated market would not provide. Although such regulation is inconsistent with a freely competitive system, the one-sidedness of competition for certain goods and services has been recognized by law; thus competition has not been enforced for this reason in the case of wages and most farm prices. With respect to some goods and services, price control has been undertaken by government agencies either to protect consumers—as for public utilities and railroads—or to protect producers and their employees—as for certain staple farm products, milk, and bituminous coal.

Control of some prices in an otherwise competitive price system cannot be assumed to leave other prices unaffected. When some prices are held above free-market levels either by private interests or by government, other prices are variously affected. A high price for class I milk leads to increased production and lower prices for other classes of milk and lower prices for butter, cheese, and other products whose prices are related to milk prices. Artificially high prices for staple farm products and restriction of output leads to an increase in output and a drop in price for uncontrolled farm

crops. Price increases for cotton by governmental manipulation stimulate the use of such seemingly unrelated items as paper towels and rayon dresses, and by discouraging exports, jeopardize the income of holders of foreign securities. In some areas, abnormally high union wage rates and restriction upon union membership tend to produce an oversupply of unskilled labor who must work at depressed wages or seek jobs elsewhere. Maintenance of high union rates in urban areas also leads to the migration of industry to less industrialized areas, and tends to raise wages in such areas. Since all prices are related, some more closely than others, the control of prices of such goods causes repercussions upon the demand, supply, and prices of many other goods. The wider the area of controlled prices, the smaller must become the role of prices in guiding production and consumption, and the greater the role of direct controls over these functions.

When virtually all prices are controlled, as during wartime, price must abdicate its usual authority in directing resources. This fact was recognized in World War II by setting up an elaborate system of controls over both production and consumption. The same type of direct controls must be used in a planned economy if resources are to be directed to other uses than the satisfaction of consumer demand. To the extent that a planned economy attempts to allow free choice to consumers and workers, it must set for itself and achieve by conscious direction the same objectives which are achieved automatically in a properly functioning, freely competitive, private-enterprise economy. Should it succeed in this, however, it would no longer be a centrally "planned" economy.

It should not be supposed that the fixing of maximum or minimum prices for individual goods or services is a novel or recent development. It is in fact atavistic, a throwback to earlier economic institutions. Price control has been practiced from earliest times. It was a part of the structure of mercantilist controls in Europe that was cast off in order for production and trade to make the enormous expansion which have characterized the past 150 years. Many people see in the trend toward peacetime regulation of industry, including much of the peacetime price control described above, a return to an outmoded and inflexible system of government control which for centuries stifled economic development. Others, however, are convinced that a competitive system cannot provide and continuously maintain an adequate standard of living for the majority and therefore argue for governmental direction of economic life.

PART V

MANAGEMENT AND CONTROL OF ECONOMIC RESOURCES

CHAPTER 24

General Aspects of Resource Control

The principal processes underlying the complex pattern of American economic life have been described in the preceding chapters. These are: production, exchange, income distribution, and consumption. As carried out in this economy they center largely on the market, on the purchase and sale of commodities and services, and on the use of money. Production provides both the commodities and services comprising the national output and the money incomes which make up the national income. The exchange process apportions this output and income among the individuals and families throughout the nation. With the income thus distributed, consumers buy the things they want, guided by the prices of these things and their personal preferences. The variegated forms of economic activity carried on by individuals, groups, business organizations, and governmental agencies are elements in the complex arrangements by which an economic system puts its resources to work to produce the commodities and incomes that permit the satisfaction of individual wants as well as public purposes.

The economic system works neither automatically nor in a vacuum. Its operations result from the countless decisions, large and small, made daily by *economic agents:* individuals, business firms, labor organizations, governmental agencies, and others. An economic agent may be described as

an entity, personal or organizational, that affects the economic process by disposing income and by using or affecting the use of economic resources. An economic agent may buy or sell something, or determine the conditions under which buying and selling is carried out; or it may produce something, or save, or lend money. Because economic agents determine the four main processes of the economy it is important to describe briefly the types of economic agents, the decisions they make, and the general framework which guides and directs their decisions. These decisions determine what shall be produced and how much shall be produced; the levels of national output, national income, and employment; the distribution of income payments and the uses to which this income is devoted; and the volume and types of consumption.

Economic agents may be classified as follows:

1. *Income receivers,* whose decisions largely regulate the disposition or use of income payments received by them from private producers and government.
2. *Business enterprise,* whose decisions set economic resources to work. This type of economic agent may be divided into
   a. *Business firms* or separate business units which perform the principal function of producing commodities and services;
   b. *Organizations* composed of separate business units which often influence the decision of their members.
3. *Suppliers of economic resources,* whose decisions influence the terms under which they are employed. This category may be divided into
   a. *Separate suppliers,* such as a person offering his services for hire or his assets for sale, rent, or loan;
   b. *Organizations,* such as labor unions, which influence or determine the decisions of the individual.
4. *Government,* whose decisions have an influence on income distribution, conditions of employment, and the level and types of output. Government as an economic agent may be distinguished as
   a. A *producer* of services and commodities;
   b. A *regulator* of many of the decisions of other economic agents;
   c. A *buyer* or consumer of part of the output of private producers;
   d. The *tax collector* whose decisions directly or indirectly appropriate to itself part of the currently produced income of the economy.

These agents operate and their decisions are effected in a context or framework of established economic relationships, law, custom, and social con-

ventions. This framework is by no means immutable; indeed some economic agents almost ceaselessly seek to modify it to their advantage just as others seek to maintain it intact. Thus the interests and activities of the several economic agents differ or actually conflict in many ways.

In a sense, income receivers guide the activities of many producers and of the suppliers of economic resources by means of their preferences and decisions in purchasing. However, not all of the activities of producers are so affected, because enterprise constantly attempts to mold consumer choices by advertising and to cultivate new wants for new products. New products are introduced only indirectly in response to known consumer demand; mainly they are introduced in expectation that new demands may be created. New products may be highly profitable, but it also often happens that resources are used for things which do not meet with eventual consumer approval. Another reason that consumer preferences do not exclusively guide the entire productive process is that government is a substantial buyer of commodities for its various purposes and perforce influences the character of productive activity. Closely related to this point is the fact that government itself is a producer and therefore an employer of resources; its productive activity in this country is not influenced materially by consumer choices because government sells only a small part of its services in the consumer market. Despite these exceptions, however, the decisions of consumers in the spending of their incomes exert a dominant influence on the character of productive activities.

The predominance of "free choice" on the part of the consumer together with a comparable freedom on the part of producers to cater to these consumer choices are fundamental features of the economies of the United States and Canada and are expressed in the function of price in equilibrating supply and demand. Likewise, the decisions of income receivers largely govern the amount saved and the amounts spent for consumption out of current income and thus bear upon the character of the productive activities of the country in a second major way. Because some consumers do not spend their entire incomes for consumption, some economic resources are saved and made available for the production of investment or capital goods and for the purposes of government. The saving decisions of income receivers therefore are a determinant of the use of economic resources.

Income receivers as consumers represent a group of *final users* of commodities and services, comprising the bulk of national output. They purchase and use up commodities and services in accordance with their preferences, purposes, or aims. Government, also, is a final user of some part of the national output (a small part except during war), and its purchases

and its productive efforts are governed by the purposes or aims—national security, internal order, the general welfare, etc.—assigned to it or assumed by it in accordance with the political practices and sentiments of the time and place. Together these final users account for the great bulk of the aggregate demand for national output,[1] and therefore exert the dominant influence on the character of the nation's economic activity. Thus the end uses to which national output is applied comprise the context for the employment of economic resources.

This general background is pertinent to a study of the decisions of those economic agents which participate in the productive process. Chapters 25-27 describe business organizations and especially the corporate form of business. The function of business enterprise is to anticipate the demands for its products and mobilize, organize, and manage other economic resources to meet these demands. Chapters 28 and 29 describe the growth and organization of labor unions. These organizations play a part in controlling the supply of labor and the conditions (e.g., wages and hours) on which labor shall be employed, and thereby seek to modify the employment decisions of enterprise in their favor. Chapter 30 discusses some of the activities of government which affect the decisions relating to the use of economic resources.

BUSINESS ENTERPRISE

Mobilization and management of economic resources in this country are primarily a responsibility of private business—(1) the individual entrepreneur, (2) the partnership, (3) the corporation, and, to a lesser extent, (4) other forms of business organization such as trade associations and cooperatives. In a partly socialized economy, such as Britain, government plays a large role in this function; in Russia, government has a virtual monopoly of the task of mobilizing and managing resources. Business organization stands between the consumer and other final buyers on the one hand, and the economic resources that (with it) are the source of production on the other hand. By the decisions of business enterprise capital is invested, human effort is employed, land and other resources are used. Intelligence, skill, and energy in mobilizing and managing these resources are of utmost importance, and therefore the vitality and performance of an economy are in large measure determined by the character and aims of its organizational and managerial talents.

[1] As described earlier, enterprise too comprises a part of the demand for national output, i.e., the demand for capital goods.

The decisions of business enterprise are motivated first and foremost by the desire to obtain a profit, for without hope of such reward business enterprise will not make the decisions that create new products, new industries, new wants, nor will it expand output of existing industries and adopt new production techniques—all essential to material progress. In the traditional conception of competitive private enterprise the earning of profit is contingent upon (1) the production of commodities and services that satisfy the wants of prospective buyers, and (2) the skillful mobilization and management of the economic resources essential to this production. The failure to gauge demand accurately or to manage resources efficiently leads to losses and possible bankruptcy. Although profit is the immediate goal, the means of its attainment in competitively organized markets is through maximizing consumer benefit. This traditional conception may be exemplified by a simple case: a family producing most of the things it wants. The head of the family presumably would manage his own time and efforts and those of other working members of his family and use his land and other resources to get the best results from these productive factors. The function of management in this simplified case is the maximizing of want satisfaction of the family.

Similarly, the individual enterpriser who owns part or all of his firm's assets and employs labor and other productive services, generally under the conditions of pure competition and imperfect competition which characterize "small business," manages these resources in terms of personal self-interest. He seeks to maximize his profit, taking into consideration his desire for leisure and his regard for the customs and conventions of the community. This end is gained only if the enterpriser gauges correctly the wants of the prospective buyers of his output, and if he manages economic resources with skill. Self-interest therefore leads the enterpriser to mobilize and manage resources (including his own time and effort) in accord with the preferences of buyers. Competition tends to force him to produce the wanted commodities and services at prices sufficient to reward him adequately for his entrepreneurial function. *Thus competition is a means of regulating the types of commodities and services produced;* ideally the composition of output suits the preferences of consumers as expressed by their purchases, the price is as low as competitive costs permit, and aggregate output is at maximum level. This is not only an explanation of the way in which the competitive economic system functions but also a social justification of the competitive free enterprise system—a social justification based on the presumed maximization of individual well-being and therefore social well-being.

In the actual operations of the economic system, this presumed harmony of the individual enterpriser's interest and the public welfare is often obscured by the complexity of the system and sometimes perverted and thwarted by it. When the management of resources, the ownership and supply of resources, and the wants to be satisfied are in large measure separate elements in a complex economic environment, want satisfaction and resource use are only indirectly related. This separation emerges when the household production unit gives way to specialized productive enterprises. Separation, moreover, is intensified when the productive enterprise takes the form of a large corporation. The function of management is then concentrated in the hands of the corporation executives. The resources they manage are procured largely through markets for labor, for capital, and for other economic resources; and the supply of these resources in turn is more and more subject to concentrated control. Labor organizations seek control over wage earners and in varying degrees influence the decisions relating to the employment of labor. The supply of funds to procure capital facilities tends to become concentrated in the hands of financial institutions.

Thus the economic resources available for the production of commodities and services are increasingly subject to the decisions made by large organizations—corporations, investment banks, and trade unions. Increasingly also these decisions are regulated by governmental agencies for a variety of purposes defined by law or assumed by the agencies of government. The very existence of these organizations raises questions as to their effect on the economy: What purposes direct the decisions made by corporate business management, by labor unions, and by governmental agencies? Is there anything inherent in large-scale organization itself which influences the decision-making power of these economic agents? Does centralization of decision-making in the hands of large organizations affect the use of economic resources and ultimately the maximum satisfaction of human wants? The economy today contrasts markedly with the traditional concept of small-scale private enterprise, a free labor market, and a governmental policy of nonintervention. To a considerable extent, these changes have brought about a shift in the power to manage economic resources from individuals to organizational leaders—corporate executives, union leaders, and governmental officials.

This is evident in the field of corporate business enterprise. In the United States today corporations are dominant in manufacturing, mining, transportation, communication, and finance, although somewhat less important in retailing and of no consequence in agriculture. In the industries where the corporate form dominates, the trend has led to the concentration of business

control in the hands of not only few but very large corporations. Two or three hundred of the largest corporations in manufacturing, finance, transportation, and communication own an enormous part of the business assets in these fields, and because of their pre-eminent position they exercise leadership over a large part of the remaining business assets. Because these large firms are so important as customers of innumerable smaller firms, the latter's decisions are indirectly bound to those of the large firms.

The growth of corporations and the trend toward concentration has undoubtedly contributed substantially to America's economic expansion. The small individual enterprise could not create a railway system, a large steel mill, or other costly parts of the productive mechanism. Large-scale operations, moreover, are sometimes more efficient than small-scale productive enterprise. The corporate form of enterprise has encouraged the growth of large and efficient productive establishments in those sectors of the economy where the optimum size of plant is large; mass production and lower cost than otherwise are features of this development in business. Mass production of automobiles, shoes, radios, cigarettes, mechanical refrigerators, telephones, and many other familiar things has contributed to higher planes of material well-being.

When profits are maximized as a result of increased and cheaper production of goods by means of large-scale enterprise, consumers are benefited. However, the decisions of corporate managers are not always directed toward this objective. The sheer maintenance of a corporation—its power, position, and prestige—is often the primary objective of management's decisions. Consumers may benefit from such preoccupation of management but not necessarily. For the urge for power may lead to monopoly in one form or another, to the possible detriment of the consumer. Expansion in size may be undertaken for its own sake, being neither profitable to the corporation nor of benefit to consumers. The cult of big business is partly sustained by the empire-building proclivities of corporate promoters; indeed some corporate systems have little other basis.

Apart from the position of the corporate entity itself, management of necessity must constantly look out for its own interests. Competition within the executive group is frequently fierce. Corporation executives must often be on guard against "outside" groups who wish to take over. The actions of management are not uninfluenced by considerations of managerial self-interest; the maintenance of position, or the improvement of position, in the corporate bureaucracy is a constant preoccupation. This is sometimes referred to as the "politics" of business bureaucracy and is similar to the internal politics of governmental and labor-union organizations.

In addition to these bureaucratic and organizational interests, there are other influences which affect management's decisions in the conduct of business. Trade associations made up of member firms in an industry and similar business organizations mold the opinion or "ethics" and sometimes even direct the actions of individual firms in matters of policy. Trade associations, such as the American Iron and Steel Institute, the Meat Institute, or the Automobile Manufacturers' Association, are common in virtually all branches of business. While some are of little consequence, others take an active leadership and frequently "represent" an industry in many matters and seek to guide the firms in the industry on some questions of business policy. Organizations of general scope such as the National Association of Manufacturers and the United States Chamber of Commerce attempt to mold business opinion and guide business decisions in such matters as labor-management relationships, wage policies, taxation, governmental control, and price policy. These organizations stress the "social responsibility" of business, its need to assume leadership on large issues of public importance, etc.

The economic decisions of business enterprise govern in large measure the types of commodities and services produced and the amount of national output. The traditional and still widely accepted view of private enterprise regards these decisions as made by the individual enterpriser who, in seeking to further his own interest, mobilizes and manages economic resources in a way calculated to produce the most at the least cost. The growth of corporations, and especially the emergence of a few hundred giant corporations, necessarily modify this conception. A substantial control over the use of economic resources is thus transferred to a relatively small group of "professional managers" employed by corporations. Their decisions tend to be influenced in some measure by organizational interests, by internal bureaucratic considerations, and by the views of organized business groups. This does not deny the efficiency of the corporate form of enterprise or the wisdom of the decisions made. It merely points to a more complex motivation underlying big business decisions when compared with the simple purposes of the individual enterpriser.

LABOR ORGANIZATIONS

The growth of labor organizations has followed, in somewhat parallel lines, the extension of big business under the corporate form of enterprise. In 1947 there were approximately 15 million wage and salary employees in unions in this country. Briefly, it may be said that union organizations

influence the decisions of both employers and employees on matters of (1) wages,* (2) hours of work, (3) the physical conditions of employment, (4) the security and tenure of the employee, and (5) the performance or amount of work done. The employment contract made by employers and employees, whether written or oral, covers these points; the terms of the contract are matters for bargaining between the parties. The strength of a union in the collective bargaining process is derived largely from the extent to which the supply of labor for a given place—firm, industry, or trade—is unionized. In some industries, such as coal mining, steel, automobile, and rubber manufacture, all or nearly all wage earners are organized; while in others, such as the service trades, there are only local parts of the total labor supply in unions. Furthermore, union control may be industry-wide or merely based on crafts, so that union bargaining with employers may involve several different situations. These may be contrasted as follows:

1. *The entire group of wage earners in an industry bargains as a unit with employers.* This is true of the coal miners, almost all of whom are members of the United Mine Workers union. The employers in turn may sometimes bargain as a unit with the single union. This also has occurred in the coal industry. Or employers may consolidate into regional groups, each group bargaining with a single union; or again one big union may bargain with each employer singly. For example, the United Automobile Workers bargains separately with each of the automobile manufacturers and negotiates a different contract with each one.
2. *Each local unit of a union bargains with local employers.* This is the practice in the construction industry. Each building-trades union in a city bargains with the several contractors of that city on terms of the employment contracts, so that each employer bargains with several company unions. A single independent local union, made up of members employed by only one employer, may bargain with him on contract terms.

The bargaining strength of unions, it has been said, depends in part upon the extent to which workers in a given industry or plant or trade are organized. In general the more completely the workers in a given employment are unionized the more effective is the union's or unions' control of the labor supply and of bargaining factors in dealing with employers. If a union embraces only a small part of a given supply, for example only 10 or 15 per cent of the employees in a plant, its effectiveness is usually limited.

This is one of the chief reasons that many unions insist on industry-wide bargaining.

Control over the supply of groups of wage earners is therefore essential to the purpose of collective bargaining. The most complete control is established by the "closed shop," which exists when an employer agrees to hire only union workers. There are many forms of preference granted by employers (usually reluctantly) to unions in the hiring, promotion, and retention of union members. Union-management agreements which give preference to union members, and legislation which gives authority to unions to act as the sole representative of workers (union and nonunion) in bargaining with employers are powerful means of inducing workmen to join and remain in unions. Undoubtedly most members regard unionization as the best means of securing increased wages and improved working conditions. The spontaneous character of early union growth is evidence of this. Membership is encouraged also by campaigns, which may merely stress the importance of union membership and the effectiveness of union actions, or may in some instances border on coercion. The motives which lead men to join unions vary, but the predominant one is self-interest—to obtain higher incomes, improved working conditions, and more equality in the bargaining process.

A union is effective as an organization if (1) it secures a wide coverage of workers in its field, (2) it secures preferential treatment and recognition by employers, and (3) if it maintains control over its members, i.e., if union leadership is authoritative. The first prerequisite has been described. Regarding the second point, under the provisions of the National Labor Relations Act of 1935 unions must be recognized and employers must bargain with them. This legislation has conferred great power on a union as a decision-maker for productive services, enabling the union (if it has the majority of workers) to bargain not only for its members but for the nonunion employees as well. It also largely accounts for the great increase in union membership over the last decade and for the growing importance of unions as spokesmen.

The growth of unions, like the growth of the large corporation, concentrates power in the hands of the executives in control, in varying degrees depending on union organization and the quality of leadership. In many unions, the principal decisions relating to wages and hours and other conditions are made by the local units of the union. Bargaining is thus decentralized. In other unions, bargaining rests largely in the hands of one leader or a small group whose decisions affect directly many thousands or hundreds of thousands of wage earners in the union. The growth of large

unions, like that of large corporations, tends increasingly to centralize the decision-making power in the hands of organizational executives or managers, either local, regional, or national. As this centralization of power has grown—in unions and in business—legislative measures have been proposed and sometimes enacted to restrain the abuses which often accompany the exercise of great power; the Sherman Antitrust Act and the Taft-Hartley Act are examples.

The individual wage earner in a union, of course, may exert an influence on policy if enough others in the union will join him. Policies have been reversed; leaders have been deposed in unions as well as in corporate enterprises. But collective bargaining inherently involves subordination of the individual to the majority; and it may involve subordination of the majority to the will of the union executive. In this sense, both unions and large corporations have traveled similar paths.

Other parallels may be noted. A union executive must produce recognizable results to maintain his position; if not, there is no dearth of ambitious rivals who aspire to his position. These results may include higher wages, shorter hours, better conditions, vacations with pay, seniority rights, health and welfare funds contributed by employers, and other advantages. Similarly the corporation executive must produce results for those to whom in the end he must answer—the stockholders. In both cases, however, power may sometimes be retained by "political means" independent of achievements in management.

The decisions of labor executives, like those of corporate executives, may not always be prompted exclusively by considerations of union member advantage. Union organizations as such develop interests. The competition between rival unions, as between the AF of L and CIO unions, or between factions within a union, is partly inspired by a desire to survive and dominate. Actions taken by one rival union may be calculated to harass the other; just as the actions of one corporation may be inspired by a "beggar-thy-neighbor" spirit. Union members in the one instance, and stockholders in the other, may derive no advantage, and may sometimes suffer a loss because of mere organizational rivalry.

These union actions do not, of course, deny the advantages of unionization to member and possibly to nonmember workers as well. The point is merely that as decision-making power shifts to an organization and its officers, the organization and its officers tend to develop their own interests while ministering to the interests of their members.

Both union and corporate executives, however, of necessity must interpret their constituents' interests to a large extent, for this is a responsibility of

leadership. In this respect, both union members and individual stockholders in a corporation are in a similar position in respect of their leaders. The individualism enshrined in tradition and in the mores of the people diminishes in importance as large segments of the nation's economic resources become subject to the management of powerful executives, both union and corporate.

GOVERNMENTAL CONTROL

The centralizing trend in the management of economic resources is accentuated and complemented by a growth in governmental intervention in economic activity. Government has entered into the economic processes as an agent in many ways; in part it has intervened in an effort to restrain or regulate the activities of organized private groups in the economy.

Some forms of intervention are *regulative,* permitting certain types of actions and forbidding others. For example, labor legislation forbids the payment of less than the legally established minimum wage and forbids interference by employers with union organization; business legislation prohibits monopolistic practices by firms when these are considered harmful to the public, and certain "unfair" competitive practices by business are likewise forbidden. Corporations are regulated in the issue of new stocks and bonds in order that these securities will not be misrepresented to the purchaser. Railroads are closely regulated by the Interstate Commerce Commission; the issuance of railroad securities, the merging of companies, the extension of new lines or the discontinuance of existing lines, changes in rates and fares—all are decisions which must be approved by the government. Banking regulations limit the different types of business that national banks may do. Radio broadcasting is subject to federal control, requiring among other things a permit for the establishment of new stations. Certain minimum standards are established in the food and drug business to prevent practices dangerous to public health. Not only these and other federal regulations, but also innumerable state, county, and city regulations work to control economic activities within their jurisdictions. For example, rates charged by street transit companies and by gas and electric utilities are subject to approval by state public-utility commissions. It will be noted that regulatory measures are largely protective; they involve the imposition of rules of conduct, especially on business organizations, which limit the decisions of enterprise in accordance with public policy.

Intervention by government takes other forms. In the field of agriculture, the government attempts to establish parity prices by purchasing part of the output of certain crops, by offering financial inducements to farmers to

restrict production, and by encouraging marketing agreements which limit the amount of produce sold. The establishment of farm credit agencies, home mortgage agencies, and the making of loans to industry tend to encourage certain types of economic activity. Government tax policies affect almost all consumers and income receivers. Heavy taxes on business income may discourage the growth of firms, while heavy taxes on personal incomes may reduce savings; excises and sales taxes tend to reduce consumption by increasing prices; and import duties tend to raise prices and encourage certain forms of domestic production. Governmental expenditures tend to encourage businesses which sell products or services to government.

Government intervention has grown tremendously with the passing decades; depression problems and war have greatly stimulated both direct and indirect governmental intervention in economic affairs. The growing complexity of the economic system, too, has encouraged control, as in the fields of banking and corporation finance. Moreover, the concentration of economic power in private groups and the tendency to abuse that power have in part fostered government control. The regulation of business became necessary as the power of business expanded just as nowadays the regulation of unions by government becomes an issue as unions increase in power.

As the regulatory activities of government expand, increasing power in economic decisions is given to the Interstate Commerce Commission, the National Labor Relations Board, the Federal Communications Commission, the Reconstruction Finance Corporation, the Commodity Credit Corporation, and other agencies. The purposes of these agencies are defined by law, and the decisions therefore are generally guided by stated public purposes. Few agencies of government, however, confine their activities to the literally expressed authorization by law. Regulation of economic activity has tended to create problems peculiar to its own establishment which then call for additional regulation and additional governmental activity not contemplated at the outset.

Organizational and personal influences are sometimes at work in governmental agencies, just as in corporate and labor bureaucracies. The position and prestige of a governmental agency grows with its size and functions, so that it is partly inspired by the urge to maintain or improve position vis-à-vis other governmental agencies. Jurisdictional union rivalry and the competitive struggle of large corporations have their analogues in government. The government bureaucracy can be a means for gaining personal advantage. Officials or executives have their own interests to protect or improve, and sometimes this takes the form of expanding the functions of the agency or affecting its official decisions.

Opposition to the trend of government control of economic decisions is persistent but unavailing for the most part, because it is usually directed against only that part of public policy objectionable to the protestant. Thus a businessman will protest a governmental action which hampers his freedom, but will support an action (such as a special tariff) which benefits him at the expense of other businesses. Labor unions embrace legislation and intervention which curb antiunion activities on the part of employers but regard any attempted legislative curb on their own actions as a plot to destroy collective bargaining. Thus, both business groups and labor unions clamorously oppose governmental intervention not to their liking but assiduously promote intervention when it benefits their interests. And both may, while promoting intervention in their own behalf, engage in monopolistic practices which merely invite restrictions repugnant to them.

CONCLUSION

In the legal, social, and economic traditions of this country, private interest operating through the market mechanism is the principal motivating influence behind the mobilization and control of economic resources. These traditions assign to competition the task of limiting and directing private interest to the public welfare. Small competitive producers serve best their own interests by catering to the preferences of buyers in the market. Competition assures that wants are satisfied at minimum costs. The competitive system is sometimes called an individualistic or "atomistic" economy; i.e., an economy characterized by large numbers of small-scale producers, a free (unorganized) labor supply, equal access to other economic resources, and governmental activity limited largely to the protection of property, persons, and public security. In such a system economic decisions that mobilize and direct the work of economic resources are highly *decentralized.* The individual consumer (or family unit) and the individual wage earner and employer each decides to spend, save, work, and produce in accordance with the individual or family conception of its own best interest. Centering about the competitive market mechanism, these decisions determine the amounts and types of investment, production, and employment in a way that maximizes the public interest.

The growth of large aggregates of private power virtually destroys the traditional conception of an "atomistic," self-regulative economy and the inherent harmony of private and public interests. Monopoly power in business led early to an expansion of governmental intervention in economic activity: to restrain monopoly growth, to prevent its development, or to

regulate its activities. Monopoly power in labor organization—partly encouraged by government—leads in similar directions. Some groups unable to organize successfully to promote their own interests, agriculture for instance, accomplish the same general purpose through political means, mainly by agricultural legislation and programs. In short, private interest tends to consolidate into organized groups exercising great economic power, or failing this, to look to government to support their interest by legislation. Private group interests supported by private organization or by public agencies pose such problems as (1) the means of reconciling group conflicts, and (2) the means of harmonizing group and public interests. Although the market mechanism resolves these problems to some extent, the market itself is increasingly subject to control by organized groups. The reconciliation of group interests and of these interests with the general interest of the economy is therefore largely in the hands of those who influence or determine the decisions made by each group.

CHAPTER 25

Business Organization; the Corporation

Private economic activity is carried on by individuals or organizations of individuals. In the final analysis, it is the countless decisions of individual persons made within the framework of law and custom, either singly or in groups, either by choice or by acquiescence in the decisions of others, that bring about the production and consumption of goods and services and the payment of incomes. At the level of final consumption decisions are made by a relatively simple economic unit, the individual household. At the level of production, decisions are made in a variety of organizational forms ranging from the individual proprietor to the gigantic corporation which controls most of the business in its field and in which hundreds or even thousands of persons participate actively or by assent in the decisions made. In general terms, the most usual forms of business organization are the proprietorship, the partnership, the corporation, and the co-operative.

Only 150 years ago it was considered unwise to engage in a competitive

enterprise if the managers of it were handling other people's money rather than their own, since self-interest in investment was held to spell the difference between success and failure. Accordingly, the corporation was thought to be a form of organization suited only for the operation of routine noncompetitive businesses, such as canals and waterworks, requiring for their operation more capital than could be supplied by a few persons. Adam Smith, for example, limited the prospect of success to corporations having either a monopoly or being engaged in such fields as banking, insurance, building and operating canals, or supplying water to cities.

The dominance of the corporate form of organization in the United States and of similar forms of organization in western Europe indicates clearly that the presumed disadvantages of the corporation have been largely offset by certain advantages, as will be shown. Table 57 shows the various proportions of total business in each industry carried out by corporations in 1937.

Table 57 IMPORTANCE OF THE CORPORATION

Industry	*Per cent of business by corporations*
Agriculture	7
Mining	96
Electric light and power and manufactured gas	100
Manufacturing	92
Contract construction	36
Transportation	89
Communication	100
Trade	58
Finance	84
Service	30
Miscellaneous	33

Source: TNEC *Hearings*, 1939, Pt. 1, p. 96.

In all, between 60 and 65 per cent of the total volume of business of the country is accounted for by corporations. They are more important than all other forms of organization in mining, manufacturing, transportation, communication, trade, finance, and government (although in the last field they are not primarily engaged in business for profit). Individual proprietorships are most common in agriculture, trade, construction, and services. The strongholds of partnerships are primarily in trade, professional service, and finance. Co-operatives are of minor importance in most fields but are relatively important in agricultural marketing and purchasing.

THE PROPRIETORSHIP

As a form of business organization most closely akin to the family unit, the proprietorship was naturally the earliest form of business organization. This may be seen in the gradual breaking down of the feudal system in the late Middle Ages (although, of course, business organizations thrived many centuries before) when trading and manufacture for trade grew more important. As serfdom was abandoned men became free to own property and engage in productive activity in their own right; thus many became individual proprietors. Similarly, the proprietorship was the most common form of business organization in America during the predominantly agrarian colonial period. It has to this day remained dominant in agriculture. Proprietorships settled the West, started most towns and cities in this country, and bequeathed to modern generations the folklore of rugged individualism which still persists despite the fact that most of the national output is produced by corporations.

The proprietorship is a form of business organization which is owned, managed, and controlled by a single individual who risks his property, and receives all the profits or bears all the loss and assumes all legal responsibilities resulting from its success or failure. It is the simplest and easiest to form of all types of enterprise. Its size and possibility of growth are dependent upon how much capital the proprietor already owns and can put into the business, how much capital he can secure by retaining earnings, and how much he can borrow. Since the business has no legal existence apart from the owner himself, in case of failure all of the owner's assets (with the exception allowed by bankruptcy laws), whether used in the business, in other activities, or by his family, can be taken to satisfy his debts. Since, in effect, the proprietor is coextensive with the proprietorship, the business ceases at his death or incapacitation. As a form of business organization, the proprietorship links the success of the business with the resources, effort, ability, and ingenuity of the person who owns it.

The proprietorship suffers severe limitations. Its duration, limited to the active life period of one person, makes it a poor instrument for long-time undertakings and generally a poor credit risk for long-term loans. Proprietorship is therefore usually limited in its ability to secure large loans. It is also limited in assuming them because, in the event of failure, business property cannot be separated from nonbusiness property in settlement of debts. The proprietorship also suffers from its apparent strength—the concentration of ownership and management in the proprietor. It is difficult

under such a scheme to obtain specialized management in other persons than the proprietor, for their fortunes are entirely dependent on the whims (and possibilities of survival) of the proprietor. For these several reasons the size of proprietorships tends to be limited to relatively small businesses. When businesses expand greatly they usually change from proprietorships to other forms.

From a social point of view, the proprietorship is considered worthy of preservation because it is a stronghold of individual responsibility and self-reliance. This advantage is often recognized by legislators, especially in dealing with such farm problems as foreclosures of farm mortgages. It is, moreover, a form of business organization which can most easily be used by new entrants into an industry not requiring a large initial investment. Industries where proprietorships predominate are likely to be highly competitive, for the enterpreneur is given full scope to foster innovation and economic progress. Finally, this form is historically the bridge over which wage workers or tenants have moved to achieve independent ownership of a business or farm.

THE PARTNERSHIP

The limited ability of proprietors to expand—to obtain large amounts of money for business, to obtain specialized managerial ability—has from earliest times resulted in the formation of partnerships. In a partnership, two or more individuals associate themselves by means of a contract in the common ownership, management, and control of a business enterprise. The enterprise is still an enterprise of persons, since the partnership ordinarily is not recognized as a separate legal entity; i.e., it cannot own things, either sue or be sued, or make contracts in its own name, but only in the names of the partners. Each partner may act for all the others, and each generally bears unlimited liability for the acts of the others. Since the existence of the partnership resides in persons, the death or withdrawal of one partner automatically dissolves the partnership.

These characteristics would seem to make the partnership subject to most of the disadvantages of proprietorships. Like the proprietorship, it is of uncertain duration. The liability is direct and unlimited, for each partner. Moreover, the partnership has the further disadvantage that agreement of two or more people must be secured upon all important actions. A partner may withdraw and thereby dissolve the partnership, but the value of the withdrawing partner's share depends in large

part upon his finding someone, possibly one of the original partners, who is willing to assume his former personal relationship with the remaining partner or partners and who is at the same time acceptable to them.

However, partnerships have certain advantages to private persons not enjoyed by proprietorships. A firm can expand and its credit standing improve by taking on the financial resources or ability of new partners. Moreover, specialization is possible in partnerships. Where special statutory authority has been granted by the states, limited forms of partnerships are possible in which one or more partners furnish financial resources but do not exercise a voice in management. Limited partners, as long as they act in accordance with the statutes, are liable only to the extent of their investment in the firm. Such partnerships can more easily secure additional money and therefore capital goods by admitting limited partners, while at the same time preserving centralized management and control. Finally, certain specialized partnerships may be used to carry out single ventures. The most common form in this country is the underwriting syndicate for floating security issues of considerable size. A number of investment-banking enterprises may join together for the specific purpose of marketing the security issue, guaranteeing a certain price to the corporation for its new issue, and undertaking jointly to sell it or to buy the unsold part of the issue themselves. Each member of the syndicate is liable for his share of the issue, and the partnership, although limited in duration to the time required to market the issue, cannot be terminated until the securities are sold. Other forms of special partnerships such as mining partnerships in the West can be formed to carry on particular business operations, for the partnership is found to be a flexible form of business organization.

Although still widely used in the modern economy the partnership suffers from its personal nature. It cannot operate as an entity apart from the persons composing it, and even in the limited partnership the share of one partner is not easily saleable. For its members, a partnership which is engaged in manufacturing or buying and selling may be regarded as risky because all of the partners are subject to unlimited liability.

THE CORPORATION

A corporation is a voluntary association of natural or legal persons, organized under and recognized by law as an artificial person separate from the persons who compose it, for the accomplishment of certain purposes specified in its charter. Its life is usually limited by its charter although it may be designated as perpetual.

The modern business corporation, while of comparatively recent origin, is related to earlier forms of organization which were invented to meet specific needs. One of these was the joint stock association or company, which was usually formed under the common law or given special privileges by statute. The Russian, East India, and Hudson's Bay companies were joint stock associations given government dispensation for the development of trade and colonization in the sixteenth and seventeenth centuries; likewise, the London and Plymouth (joint-stock) Companies founded Virginia and Massachusetts. Joint-stock companies were especially sponsored by law because they were a practical means by which to obtain large financial resources and centralized management necessary for large-scale foreign trade and colonization. Joint-stock companies developed out of the partnership in that their owners were subject to unlimited liability, but they resembled the modern corporate form in that the owners received shares in proportion to their investment which could be freely traded.

The modern business corporation also owes its origin in part to the charters granted to towns by state rulers in the Middle Ages. By granting a charter the ruler delegated power to others; his interests were managed by appointees. This practice affected commercial development (as well as political) for it impressed upon owners the means of separating ownership from management by the delegation of powers. The corporation today is in this sense granted a charter to conduct business with the resources of its owners.

Perhaps the most compelling reason for the use of the corporate form of organization was the advent of the industrial revolution. New and improved technology made profitable the use of large amounts of capital in combination with other resources. It was soon recognized that a form of business organization which could attract the large amounts of financial resources required for the new factories without unduly cramping the powers of entrepreneurship was necessary to facilitate and encourage industrial expansion. The corporate form of business organization filled this need, but its use has long since been extended to other enterprises not specifically requiring it for capital raising.

Unlike a proprietorship or a general partnership, a corporation in the United States can be formed only after authority has been granted by the state (or sometimes the Federal) government. Until about a century ago, corporations could be organized only by securing a special law from the state legislature or from Congress; the corporate charter was regarded as a special privilege and was accordingly granted only sparingly. During the last century, however, the granting of charters has been progressively liberal-

ized, until states now compete with one another to attract the business of incorporation and so secure the revenue from the franchise tax and certain other taxes levied on corporations. By the 1860's most states had enacted general incorporation laws under which charters could be granted without special legislative action, and this is now generally the case. Delaware, Maryland, Maine, South Dakota, and Nevada prescribe a minimum of restrictions. Lawyers in these states have long been able to offer by mail the service of delivering all the documents necessary to organize a corporation to operate anywhere in the United States. Corporations organized under the laws of one state may do business in any other state provided they meet its requirements for "foreign" corporations.

Corporations do business under a state charter which at least nominally limits the corporation to certain lines of business, although the purposes of the corporation are usually stated in such general terms that there is but little restriction. The charter of the United States Steel Corporation, for example, seems only to exclude it from the fields of banking and insurance. Activities not authorized by the charter of the corporation or the law under which it was established are *ultra vires;* its officers may be subject to penalties for undertaking them, or the charter may be revoked if it is proved the stockholders have authorized the activities.

The ownership of the corporation is vested in the stockholders, each of the original stockholders being issued a number of shares divided by agreement, usually in proportion to investment in the business or the value of services rendered in connection with the incorporation. Shares are freely transferable; the owners of a corporation can be and often are a constantly changing group of persons. The shares representing ownership may be of one or several types [1] depending upon how the corporation has sought to obtain capital. Owners of the corporation, i.e., of the shares of stock, are usually subject only to limited liability, their possible loss in the event of failure of the business being limited almost always to the amount they have paid (or agreed to pay) for their shares of stock.

A corporation is controlled by those stockholders—usually the holders of one class of stock—having the right to vote for directors. The principle commonly followed is one vote per share, although in some corporations stockholders are allowed a number of votes equal to the number of shares multiplied by the number of directors to be elected. The working rules of the corporation are the bylaws, which are subject to change usually only after approval by stockholders. Bylaws specify how stock is to be issued and transferred, how dividends shall be paid, when direc-

[1] See Chap. 26, pp. 563-64.

tors' meetings are to be held and in what manner, the qualifications of directors, the number and duties of officers, and similar matters of procedure. The actual management of the corporation is left to the officers, who in many cases are substantial owners as well. In the smaller corporations ownership and management may reside in the same persons. In large corporations, however, ownership and management are usually separated, and the powers of management are consequently great because the large number of stockholders can only with difficulty exercise their rights. Management, by obtaining proxies of stockholders, can and often does perpetuate itself in office. A proxy is a power of attorney giving to someone else the right to exercise the stockholder's vote. So long as the existing management can keep the stockholders satisfied—usually by paying dividends with reasonable regularity—stockholders permit it to continue in office by granting proxies to management.

Perhaps the greatest advantage of the corporation as a form of business organization is its ability to raise money, i.e., to attract capital and to expand its facilities. Stockholders are, except in rare cases, liable for the debts of the corporation only to the extent of their investment; they are free to dispose of their shares at any time without terminating the life of the corporation or necessarily impairing its value to potential owners of the shares to be sold. Investors may usually buy corporate securities suitable to their own tastes and interests, some securities showing an emphasis upon security and regularity of income, others bearing considerable risk of loss but also chance for large gains. If the corporation is successful and wishes to expand it may issue (in some instances after amending its charter) new securities to obtain the necessary funds. If it wishes to absorb a competitor or to take in another firm making other products, it may issue its own securities in exchange for the stock of these companies.

It is the enormous capacity for growth or expansion for the purpose of obtaining economies of operation, of realizing the profits of monopoly, of satisfying the desire for power, or for other reasons, that has made the corporation the target of many attacks.

The pessimistic prospects for corporations set forth by Adam Smith and earlier economists have not been borne out in practice. Modern industry in this country and in western Europe probably would have been impossible without the aid rendered by the corporate form in amassing the capital required for large-scale production. Control over details in organization has been achieved by such devices as careful record-keeping, independent audits and reports to stockholders, piecework wages, bonuses to supervisory personnel, and promotions for efficiency. Management has been impelled

to exercise the required initiative and flexibility by the spur of competition and the necessity for making profits to retain control, and has been enabled to do so as a result of receiving authority from inactive stockholders. The corporation as a form of organization has succeeded so well, in fact, that it is singled out for special taxation by state and Federal governments. The federal corporate income tax, which does not apply to other forms of organization, has in recent years taken a considerable proportion of corporate profits, despite the fact that stockholders are subject to a personal income tax on the corporate dividends derived from these profits.

The Holding Company

The holding company is a latter-day outgrowth of the corporate form of organization which has in most recent times been the object of considerable Congressional investigation and reform legislation. The holding company is a corporation formed or used primarily to acquire the voting securities of other corporations and thereby to control them. Formerly illegal under common law, its general use began with the revision of the corporation laws of New Jersey in 1889, followed by other states' amendments. This form was given impetus when early companies like Standard Oil and American Sugar Refining companies sought to secure monopoly power by retaining or acquiring control over competitors through buying their stock.

There are several reasons for the use of holding companies, although some of these are not generally held to be socially desirable. First, a holding company may be utilized to secure a consolidation of two or more companies. A parent company merely buys up or exchanges for its own stock a controlling interest in the stock of the other company or companies. Control can be secured by this method with a much smaller expenditure than would be required if the parent company directly bought out the other company or companies. Although this use of a holding company has probably made the establishment of monopoly power too easy, outlawing its use would not actually prevent consolidation or merger of companies since this can be achieved by acquisition of one company's entire assets by another.

Second, a holding company is usually able to attract capital to subsidiary firms on better terms than they could obtain for themselves. Many of the great public-utility holding companies were favored by investment bankers since a large holding company is generally considered safer and is usually better known than the individual companies it controls. An example is the American Telephone and Telegraph Company which is

a holding company for a considerable proportion of the telephone operating companies and for research and equipment-making units, including the Western Electric Company. Other examples—notorious for their callous disregard of stockholders—were the Associated Gas and Electric and Insull utilities systems which were involved in financial scandals during the 1930's. The major defect of the holding company as a device for raising money cheaply is the possibility, often exploited in the past, that the economies obtained in this way will not be passed on to consumers or investors but rather will be siphoned off by and for the controlling group in the holding company.

Third, the holding-company form of organization has also been used to avoid legal difficulties. Some states may have onerous restrictions on "foreign" corporations which can be avoided by forming a separate domestic corporation in each state, all of them owned by a single holding company. While it serves to overcome the excessively narrow sectionalism of some states, this device unfortunately also removes from state supervision certain undesirable corporate activities. In many cases, the use of the holding company has created a "twilight zone" in which neither state nor federal control is exercised. Another legal difficulty avoided by the use of the holding company is the preservation of the business name and good will of a well-established firm. To all outward appearances the company remains the same despite holding-company ownership, whereas an outright merger would cause the company to lose its identity.

Finally, probably the principal reason for the holding company device is that it permits a maximum of control with a minimum expenditure in the purchase of stock, by means of manipulating certain corporate securities. Briefly, corporate securities may be classified into three general types: (1) *common stock*, which generally carries the right to vote for directors and so to control the corporation; (2) *preferred stock*, which usually carries no voting rights except in very unusual circumstances, and (3) *bonds*, which are long-term loan contracts and have no voting power whatsoever under ordinary circumstances. Bonds and preferred stock are given certain features such as a fixed rate of return and a prior claim upon earnings which make them attractive to more conservative investors (as will be shown in Chapter 26).

Control over enormous aggregations of corporate wealth can be secured by constructing several layers of holding companies, each financed by two or three of the types of securities. Suppose that three operating corporations each have securities outstanding as in Table 58.

Table 58 SECURITIES OF THREE CORPORATIONS

	Company A	Company B	Company C
Common stock	$ 5,000,000	$15,000,000	$ 5,000,000
Preferred stock	2,500,000	10,000,000	5,000,000
Bonds	2,500,000	5,000,000	None
Total*	**10,000,000**	**30,000,000**	**10,000,000**

* Security values are assumed to be current market values.

Since only common stock has the power to control, ownership of anything over 12½ million dollars' worth of common stock will control the three corporations having outstanding securities of 50 million dollars (and assets of perhaps two or three times that amount). Now suppose that holding company X is formed to hold a total of 13 million dollars in the common stocks of companies A, B, and C. It succeeds in selling its own securities on the following basis:

Holding Company X

Common stock	$5,000,000
Preferred stock	5,000,000
Bonds	3,000,000

The amount of stock needed to secure control of companies A, B, and C is now reduced to anything over 2½ million dollars.

If the promoters of company X now organize company Y and issue 1½ million dollars in preferred stock and 1 million dollars in common stock, they will need only a little over $500,000 to control the whole structure. If another layer is piled on, the financial resources needed for control can be cut to perhaps $200,000. Fantastic as this process may appear it has actually been accomplished by devices which, although they were more complicated, conformed in principle to what is here described. In 1930, the Van Sweringen brothers controlled over 2 billion dollars of railroad assets with less than 20 million dollars, and more than a decade later their successor bought their railroad empire at a bankruptcy sale for a fraction of this amount by buying control of the top holding company. It should be noted especially that holding companies built upon the common stock of operating companies are extremely vulnerable to fluctuations in the earnings of operating companies. If the latter do not pay dividends, the holding company receives no revenue since it has no income from productive operations of its own.

Special Regulation of Holding Companies

In the gas, electric, and railroad industries, the holding-company device was abused by some groups and caused after 1929 such large losses to investors that it was subjected to special regulation. Control was frequently pyramided in such aggregations as the Associated Gas and Electric Company. Operating companies controlled by the holding company were sometimes "milked" by being overcharged for management or construction services provided or contracted out by the holding company, by the sale of securities at inflated prices, and by loans at higher than competitive rates made to operating companies by the holding company, as well as by other similar devices. Since the holding company controls the operating companies, any transaction between parent and subsidiary is not a free market transaction unless outsiders are allowed by the holding company to bid for the sale or purchase. These abuses of holding-company power during the 1920's reached the proportions of a national scandal at a time when the stock-market crash of 1929 and subsequent corporate failures prompted public investigations. Large numbers of stockholders in holding companies saw their lifetime savings vanish when operating companies could no longer pay dividends or were forced to cut their dividend rates.

After extensive investigations, Congress passed the Public Utility Holding Company Act of 1935 to regulate and control holding companies in the gas and electric power industries. Holding companies in these industries were required to register with the Securities and Exchange Commission (SEC), which had recently been set up to supervise corporate financial practices generally and the security markets in particular. Registration consisted of filing complete information about the holding company and its related subsidiaries or associated companies. Future security issues could be made only with the approval of the Commission, which was charged with protecting the investor's interest. Future acquisitions of new companies or extension of control by holding companies were made subject to review to determine whether they were economically and financially justified. Of great importance is the Commission's power to review existing holding-company systems and thereby simplify financial structures when necessary. All holding companies above the second layer are prohibited by this legislation. In addition, companies are restricted to the control of those properties necessary to form a geographically integrated public-utility system. Other features of the law are strict prohibitions against political contributions by these public-utility holding companies,

requirement of registration of their lobbyists, and publication of other activities previously kept confidential. The SEC has had considerable success under the Act except in the matter of requiring simplification of the structure of the larger holding companies, a process which is still in progress at this writing.

Some Corporation Problems

It has been stressed that one of the characteristics which, for private purposes, gives the corporate form the greatest advantage over other forms of business organization is its ability to attract capital. This private advantage of the corporation may become something of a disadvantage to the public. Corporations have been able to grow to gigantic size which may be a threat to socially desirable purposes.

An indication of the size of America's corporate giants is the proportion of the assets of all corporations (excluding banking and other financial institutions) controlled by the largest companies. In 1929 the 200 largest corporations controlled almost one half of all nonfinancial corporate assets; and the control over the remaining half was dispersed among over 300,000 corporations. The proportion of total assets controlled by the 200 largest corporations increased during the depression years 1929-33, as shown in Table 59, but was probably reduced in subsequent years. For example, concentration of assets in the largest manufacturing companies was somewhat less in 1946 than in 1939.

Table 59 **ASSETS OF NONFINANCIAL CORPORATIONS**

	1929	1931	1933
200 largest	$ 98,597 millions	$101,662 millions	$95,617 millions
All others	100,832	76,766	72,104
Per cent of total controlled by 200 largest	49.4%	57.0%	57.0%

Source: National Resources Committee, *The Structure of the American Economy,* Washington, 1939, Pt. I, p. 107.

It has been estimated that in 1929 the assets of the 200 largest nonfinancial corporations comprised about 22 per cent of the national wealth. The effects of this concentration on the total economy are manifold. Where a few corporations dominate an industry, prices are subject to oligopolistic influences.

The consequences of this market condition on prices and output were discussed in Chapters 15 and 16.

The concentration of corporate wealth places great influence in the hands of a relatively small group of men. The 200 largest nonfinancial corporations had approximately 2,000 directors in 1930 but of these a large proportion were inactive.[2] Thus only a few hundred active directors actually presided over the declaration of dividends, decisions to make new investments in capital facilities, and other vital economic matters affecting half the nonfinancial corporate wealth of the country. If these men are generally optimistic and approve programs of capital expansion, their decisions to increase investment and thereby employment will materially influence national output; if they are doubtful or pessimistic, their decisions to retrench may reduce national output. It is significant that during World War II there was organized the Committee for Economic Development composed of businessmen who sought to lay plans for the maintenance of employment in the postwar period. If the large corporations act in keeping with the recognition of their own influence, they can become powerful instruments for maintaining prosperity; if, however, they act with respect to investment and employment as though their own actions were of little consequence to the rest of the economy, then the large corporations are likely to increase the difficulty of maintaining stable levels of output and employment.

Another problem related to the size of the largest corporations is the partial separation of ownership from control. Nominally, of course, the owners of the voting stock of a corporation also control it. Actually, however, in most of the larger corporations ownership is so widely diffused that control is possible in many cases through ownership of a relatively small fraction of the total stock. Some impression as to the diffusion of ownership can be gained from the fact that over half the stock issues of the largest corporations have been to between 10,000 and 100,000 stockholders. It has been estimated that nearly half the shareholdings of issues of the 200 largest corporations amount to $500 or less; 95 per cent amount to less than $5,000. It follows that the majority of people whose ownership is represented in one of the corporate giants have only a small stake in it. One of the more than 600,000 shareholders of American Telephone and Telegraph Company who owns only a few hundred dollars' worth of stock is not likely to regard the corporation's affairs as a matter of much concern so long as he gets his dividend regularly.

[2] A. A. Berle and G. C. Means, *The Modern Corporation and Private Property*, 1932, Macmillan, p. 46.

In contrast to the diffused and small interests of the majority of shareholders there is, of course, the very concentrated interest of those who own substantial blocks of stock or who actively manage the companies. It should be understood that a majority of the voting shares is necessary for control only in the case of an ultimate and crucial test of strength between factions among stockholders. The majority of a quorum attending an annual meeting is sufficient for working control, and a quorum need not consist of a majority of stockholders or a majority of the voting shares. It is possible for a minority interest of as little as 10 per cent of the total shares to control a very large corporation successfully for many years. If the record of the minority in control is good, it can, when a test of strength arises, solicit proxies to obtain the votes of a majority of the shares. Often, however, minority control may lead to damage which cannot be undone even when the majority seeks to assert itself. The minority may be represented in lucrative official positions for which large salaries and bonuses are paid; or it may, as occurred sometimes in the past, drain the profits of the corporation by placing contracts at extravagant prices with other companies owned more or less exclusively by the minority interest. If the minority control is exercised by the management, it may lead to the continuance of a losing activity long after the corporation's business should have been liquidated for the benefit of the stockholders, or in some cases, it may lead to profitless expansion to serve the personal vanity of corporate empire builders.

The principal methods of achieving minority control, which is usual in most of the largest corporations, are relatively simple. First in importance, of course, is the almost inescapable apathy of the average shareholder. This coupled with the fact that the incumbent management holds most of the instruments needed to secure control—stockholder lists, knowledge of the true condition of the company, and means of preventing others from obtaining access to such knowledge—sets the stage for minority control. The legal devices which may be used under the circumstances are numerous. Certain classes of stock can be given more than one vote, and multiple-voting stock can then be concentrated in a few hands. Before 1929 some corporations even went so far as to issue nonvoting common stock which promised nothing and gave the holders of the stock no rights. Or a controlling stock interest can be turned over to trustees—a voting trust—to be exercised by the trustees in the interest of the minority. Finally, a holding company or a pyramid of holding companies may be used.

CO-OPERATIVES

Although it was neither the first co-operative nor the first enterprise to operate upon co-operative principles, the co-operative store opened in 1844 by the Society of Equitable Pioneers in Toad Lane in Rochdale, England, marked the real beginning of the consumers' co-operative movement as it is now known. The store was founded by a group of weavers for the purposes of saving, in the prices they paid for their food and clothing, what they believed to be the excessive profits and costs of production and distribution charged by private distributors. It also had the object, still important to the more self-conscious consumer co-operatives, of laying the foundation for a society in which co-operative production and distribution would eventually displace private property and competition. The more important farm co-operatives, insurance companies, and similar types of co-operatives in this country, however, have no such purpose, but are superimposed on the existing structure of private enterprise and are designed to supplement and improve it.

The Rochdale Pioneers organized and operated their first retail store on the basis of the following rules, which have since become standard for co-operative organizations:

1. Each member, regardless of his financial contribution, is limited to one vote, and members control the organization.
2. Profits are distributed in proportion to purchases of the members.
3. A fixed rate of dividend is paid on the money paid in as capital by the members.
4. Sales are made at competitive prices, savings being passed on to members in the forms of dividends based upon volume of purchases.
5. Membership is open upon equal terms to all persons who have the same status as existing members, i.e., in consumer co-operatives, to all consumers; in marketing co-operatives, to all sellers of the commodity being sold.

In principle co-operatives are controlled by a one-man one-vote system, in contrast to the corporation where voting power is proportional to ownership of shares. Profits of the co-operative are distributed in proportion to the patronage which is presumed to have made them possible, whereas in a corporation they are distributed roughly in proportion to the amount of capital invested. A co-operative is therefore a form of business organization which is owned by the persons who use its services,

which is controlled in principle by all of them equally, and which distributes its profits to the members in proportion to the use they have made of its services. In large co-operatives, however, management and ownership tend to become separated, as in the large corporation. The members of a co-operative suffer no less than the stockholders of a corporation from a disposition to become apathetic; and self-perpetuation of management is equally evident in both forms of organization.

In most states, co-operatives may be incorporated under a general law in much the same manner as business corporations. Under such laws, members are subject to limited liability, but unlike stockholders in corporations they are limited to one vote. If co-operatives are unincorporated, the members are individually and jointly liable for the debts and acts of the co-operative association. Co-operatives may make contracts with their members; thus in the agricultural marketing co-operatives members usually contract to sell all their production through the co-operative.

Co-operatives may raise funds by a number of methods but are far more restricted in such activity than private corporations. Initial financial resources are raised by the contributions of members, a minimum payment for membership being the general practice. Membership shares in many co-operatives are not freely marketable, as are the shares of stock in corporations. For additional funds, the co-operative is limited to loans from present or new members and outsiders and to reinvested earnings. Their limited ability to raise money has imposed a serious restriction upon growth, which has been removed in the case of farm co-operatives only by federal credit agencies.

Co-operatives exist in many fields besides retail trade; they operate successfully in insurance, consumer and farm credit, housing (where building and loan associations are well known), wholesaling, public utilities, and —especially in Europe—in production. In the United States, consumer co-operatives have enjoyed a long history and considerable growth but have remained relatively insignificant in terms of proportion of the total volume of business. The marked difference between the American and the European experience is attributed to many factors. One of the most important is the higher level of individual income and of living in the United States; this fact has lessened the apparent importance of whatever savings a consumer co-operative may provide its patrons. Another is the competition in mass distribution through chain stores, department stores, and mail-order houses. Consumers in this country have not had to rely upon co-operative competition to force distributive margins down. Another important reason for their relative insignificance in the total economy has

been the absence of organized consumer interest in co-operative activities, particularly among wage earners. Workers and their unions have concentrated upon the more tangible benefits of collective bargaining over wages, although they exhibited a renewed interest in co-operatives when the cost of living jumped after World War II. In addition, co-operative organizations have suffered from poor management, a lack of understanding of co-operative principles, and the active opposition of established private businesses.

In farm marketing and purchasing, however, co-operatives have demonstrated considerable strength in the United States. Farmers' marketing co-operatives, including 2,730,000 members, did 4.4 billion dollars of business and farmers' purchasing co-operatives, including 1,520,000 members, did almost 750 million dollars of business in the 1943-44 marketing season.[3] Marketing associations accounted for more than three quarters of the farm co-operative business in the period given here. Co-operative marketing organizations have been authorized and fostered by government to improve the bargaining position of farmers. Co-operative purchasing of fertilizer, feed, tools, implements, petroleum products, and other farm supplies has also grown rapidly in recent years. In fact, co-operatives have reached backward to control production. Several now operate oil refineries, machinery plants, feed mills, and lumber mills. Spurred on by loans of up to ½ billion dollars from the Rural Electrification Administration, co-operatives have also organized power systems. They have contributed, along with private companies, to the expansion of electrical service to farms—from under 11 per cent of the total number of farms in 1935 to almost 40 per cent in 1942.

Another major field for farm co-operatives is farm credit. Long-term credit—for buying land, buildings, etc.—is provided for farmers by the Federal Land Banks through National Farm Loan Associations, and short-term loans—for buying machinery, stock, etc.—is provided by Production Credit Associations. Both types of associations are sponsored and financed by the Federal government, which also maintains banks for co-operatives to finance the various farm co-operative marketing, purchasing, and service enterprises. The existence of these many federal financial aids and the large number of government workers who render specialized assistance to farm co-operatives casts considerable doubt upon the indigenous and spontaneous nature of the farm co-operative movement. It seems certain that without

[3] House Committee on Small Business, "The Competition of Cooperatives with Other Forms of Business Enterprise," House Report No. 1888, 79th Congress, 2nd Session, 1946.

extensive government aid the growth of the various farm co-operatives would have been considerably slower and more difficult.

An added element of strength of the co-operative form of business organization is its favorable tax status. Whereas the earnings of corporations are subject to double taxation by the Federal government, once by the corporate income tax and a second time when they are paid to individuals in the form of dividends or interest, the income of co-operatives is taxable but once, and then only when it is paid to investors as interest and, in the case of producer co-operatives, to members as patronage dividends. So long as corporate income taxes are relatively high and so long as co-operatives are exempt from paying them, there will be some incentive to form co-operatives in order to avoid taxes. The major discrimination against corporations involved in exempting co-operatives from corporate income-tax liability is the power given to co-operatives to expand out of tax-free retained earnings held in trust for members.

CHAPTER 26

Business Finance

The prosperous business enterprise constitutes a source of wealth and income beneficial not alone to its promoters and owners but to the entire economy as well; it increases the wealth available to the community while at the same time providing investors with incomes. Similarly, the unsuccessful new venture may be considered to injure the whole economy by dissipating wealth already accumulated. Economic advancement necessarily requires the elimination of both inefficient and decadent firms; the temporary loss to the community of their failure is an inevitable price of progress.

Capital has been called the life substance of a business enterprise.[1] Original capital, obtained from persons willing to commit funds for a long term, is required in a new venture to pay costs of promotion, to purchase

[1] The term *capital* is employed in this chapter in its customary business and accounting definitions. In earlier chapters capital referred to (physical) capital *facilities*, and other capital goods.

property and equipment, and to provide in advance for initial operating expenses. Then, once the business has begun to function, additional capital is required for the firm's growth and expansion as well as to meet unexpected contingencies. These additional funds may be secured through retention of business earnings, short-term borrowings, and extension of credit by suppliers and wage earners. To continue prosperous, a business enterprise must have adequate capital to retain control of its productive agents—capital facilities, materials, and labor—and to prevent claims by its creditors from materially hampering operations or weakening confidence in the firm. Sound financing, coupled with proper financial management, is a requisite to success of any business venture.

Methods of financing and the amount of capital available to a business depend on its nature and the form of its organization. For an unincorporated firm the sources and amount of available capital are limited by its necessarily uncertain life and the fact that the assets of the firm and those of its owner are inseparable. In general, unincorporated firms find their principal source of available outside capital in bank loans, the amount being restricted to the limit of the personal credit standing of the proprietor or partners. On the other hand, a corporation has access to various sources of capital in generally larger amounts. As a legal entity separate and apart from its members, the corporation is not limited in longevity by the life span of its members; its power to attract capital may be greater than that of its combined owners. Since the corporate form of organization is the most important in the modern economy, and since the principles of corporate financing comprehend those of other types of organization, the analysis of business finance logically centers on the financial problems of the corporation.

FINANCIAL REQUIREMENTS OF CORPORATIONS

The total capital needs of a business may be classified under two main headings—fixed and circulating—although there is not always a well-defined line of demarcation between the two. *Fixed capital* is the corporation's investment in permanent or semipermanent capital facilities such as land, buildings, machinery, and equipment; it is capital tied up in specialized form for a number of years. *Circulating capital* is its investment in inventories, receivables, and money or its equivalent; it is tied up for a relatively short period of time, usually less than one year. The term *circulating* derives from the circular flow of capital, in the course of normal business

operations, from inventories to receivables [2] to money and back again to inventories. If an item is expected to be converted into money within the space of one year, it is customarily regarded as circulating capital.

The total amount of capital required, and the desirable ratio of fixed to circulating capital, vary greatly with the type of business. It is necessary for an electric power company to maintain a relatively large total investment, mostly in fixed plant and equipment, while on the other hand the total investment of a trading company may be relatively small and consist almost wholly of circulating items. Although methods of financing are different for two such enterprises, the capital invested in both is to a degree similar because in most companies a portion of the circulating capital is in actuality fixed. If the trading company maintains an inventory of merchandise and sells on credit, to operate successfully it must keep both stock on hand at all times and a cash reserve for current expenditures. Moreover, it will always have some amount of investment in accounts receivable covering recent sales. This irreducible minimum investment in inventories, receivables, and cash, although not separable from other circulating capital, is in effect just as permanent as the fixed capital. As a business grows the amount of this "fixed" circulating capital usually increases.

There is no simple formula by which the capital needs of a corporation can be precisely determined, but certain governing factors must be taken into account. Fixed capital requirements are estimated on the basis of the expected volume of business and the cost of the plant and equipment needed to produce that volume. In evaluating a corporation's circulating capital needs, important considerations are the character of the business, the length of processing period or merchandising cycle, the rapidity of turnover, scale of operations, volume and terms of purchase and sale, and cyclical and seasonal requirements. Consider the following examples in this connection. A city transportation company, operating on a money, pay-in-advance basis, requires relatively less circulating capital than a merchandising company with a large investment in inventories and selling on credit. A manufacturing concern, because of the time length of the manufacturing process, requires a large amount of circulating capital because of the size of its materials and goods-in-process inventories. (The longer the cycle of manufacturing or merchandising and the slower the turnover, the greater are the capital needs.) In general, circulating capital needs vary directly with the scale of operations and length of time through which credit is extended on sales; and it varies inversely with the credit period on purchases. The firm with a seasonal business is confronted with the necessity of financing peak in-

[2] When a firm sells on credit the amount due is a receivable.

ventories, as illustrated by the heavy investment of department stores in merchandise for the Christmas season.

Whatever the type of enterprise, provision must be made for ample *working capital,* which is the excess of circulating capital over short-term liabilities—amounts owed to suppliers for materials, to banks for short-term loans, etc. Ample working capital is needed to preserve the credit standing of the firm and to provide for contingencies such as a business depression.

SOURCES OF FINANCING FOR CORPORATIONS

Stated broadly, there are three sources of capital available to a corporation: owners, creditors, and earnings. Owners contribute their capital without any commitment—stated or implied—that it will be repaid as long as the corporation has use for it. Such capital is termed *equity capital.* But creditors make their capital available on condition that it will be returned at some specified date, and such capital constitutes a liability of the corporation whether obtained by means of long-term bonds, short-term borrowings, or by charge-account credit extended by suppliers (open accounts payable). Unless the capital supplied by the ownership is sufficient to form a secure base for the capital structure, creditors' capital will not be attracted because creditors want security. After a corporation has begun to show profits (after making payments to meet fixed obligations to creditors) it may elect to increase its capital by retaining all or part of the residual earnings rather than disbursing them as dividends. This method of financing is frequently used in order gradually to strengthen or expand the company.

Owners

The first step taken by promoters in the actual organization of a corporation is to obtain a charter from the state of its legal residence. The charter authorizes the issuance of capital stock to the contributors of ownership capital and specifies the types of stock which may be issued, the number of shares of each type, and regulations governing issuance and disposition. In return for his contribution, an individual owner receives a stock certificate, representing his equity (ownership interest) in the business and indicating the number of shares purchased. Shares of stock are transferable; an owner may sell or otherwise dispose of his equity at any time. In addition to his individual contract with the corporation represented by his stock certificate, each shareholder is also bound by the underlying stock contract

which exists between the corporation and the collective group of stockholders—namely, the body of corporate law under which the charter is granted, the charter itself, and the bylaws of the corporation.

Common Stock

The stock contract specifies the extent and manner of the stockholder's participation in the several aspects of ownership—control, income, and risk. The holder of common stock possesses (1) the right to control corporate activities, (2) the right to receive all profits in excess of fixed obligations payable to others, and (3) the burden of risk accompanying residual claims to the extent of his investment in the corporation. The management function is, in general, vested in the common stockholders (proportional to the amount of voting stock owned) who exercise this function by electing a board of directors. Balanced against this privilege is the great financial risk of investment in common stock. The fortunes of the common stockholder are closely bound up with the success or failure of the company; he stands to gain or lose heavily. The common stockholder's claims on both earnings and principal are residual to those of creditors and preferred stockholders, which means that he is entitled to any and all earnings remaining after fixed obligations have been met, or, in case of liquidation, that he is entitled to property remaining after creditors and preferred stockholders are satisfied. There is no fixed dividend rate on common stock, and dividends on this stock can be considered a liability of the company only after they have been declared by the board of directors, who are at liberty to weigh the company's need for funds and can pass (i.e., fail to pay) a dividend without placing the corporation in default. Thus, in periods of prosperity, the common stockholder stands a chance to receive a large return on his stock, far above that of the bondholder or preferred stockholder, while in periods of depression he is likely to receive nothing at all.

Certain special types of stock are often classified as common although they may include specified preferences not generally associated with common stock or may waive certain of the usual common-stock rights; e.g., the inclusion of a fixed dividend rate or the abrogation of voting power. The term *common stock* seems almost a misnomer for such issues, for they do not represent true venture capital. Since a certain amount of true venture capital is a requisite in any business and since a portion of the stock must be of the regular type, these special stocks must be regarded as supplementary issues.

Preferred Stock

Preferred stock embodies ownership, but by virtue of modifications in the contract of issuance it is endowed with preferences or limitations which differentiate it from common stock. Ordinary preferred stock has no voice in management but it does have a prior claim on earnings and property as well as a fixed dividend rate. Although preferred stock dividends like those on common stock become a liability only after they have been declared, they must be paid in full before any dividends can be paid on common stock. Most preferred stock has a stated value—usually par—which is a preferred claim against the property and which in case of liquidation of the business must be paid ahead of any distribution to common stockholders. Unlike common stock contracts, there is often a provision in the preferred stock contract permitting the corporation at its option to retire ordinary preferred stock upon payment of a stipulated amount, usually slightly in excess of par. Preferred stock may of course be bought and sold in the market like common stock.

Of several special types of preferred stock, cumulative preferred and participating preferred are especially important. Cumulative preferred stock requires the payment of all dividends which have been omitted before any division of earnings can be made among stocks of lesser priority. In the absence of the cumulative feature the corporation may pass a preferred dividend without incurring any liability for its future payment, but such omission precludes any distribution of earnings to common stock in that period. The second class, participating preferred stock, shares with common stock any dividends to be paid in excess of the specified rate on preferred plus a stipulated rate on common.

Creditors

The financial needs of a corporation can seldom be satisfied by owners' capital alone. Practically all corporations rely to some extent, or may even be largely financed, on borrowings. Such creditorship capital may consist of long- or short-term loans, secured or unsecured, or merely open accounts payable. Creditorship capital differs essentially from owners' capital in that it carries a definite obligation of the corporation to repay the principal of the loan at some specified date or according to a definite schedule, and in many instances also to make specified periodic interest payments to the creditors. Most long-term borrowings, and some of shorter duration, are

effected through the issuance of bonds to the creditors. By this technique of borrowing, securities which may be bought and sold in the market are issued in exchange for creditors' capital, and because this method permits a large number of persons to participate in the same loan, the corporation is thus enabled to borrow a larger amount of capital than it could secure from a single lender or a small group.

Bonds

Bonds are evidences of creditorship capital issued under a contract called the indenture. This contract is made with a trustee who represents the interests of all the bondholders. The indenture is a formal document containing provisions deemed necessary to protect interests of both the corporation and bondholders. The principal provisions of a bond indenture may be summarized as follows:

1. Description of the entire bond issue, including amount, interest rate, date of payment of principal and interest, denomination of bonds, type of bonds, and provisions, if any, for advance payment of principal.
2. Descriptions of the character and extent of property against which bondholders may levy to satisfy claims against corporation.
3. Special protective agreements (covenants) to ensure preservation of property pledged by the corporation as security for the issue.
4. Definitions of actions which bondholders may take to protect their interests in case the corporation fails to fulfill its promises.
5. Statements covering duties, qualifications, and responsibilities of the trustee.
6. Covenant on part of bondholders through the trustee that liens against corporate property will cease and the corporation be no longer bound by its promises made in the indenture when all bondholders' claims have been satisfied.

Bond issues vary in terms of security, term, interest rate, methods of retirement, payment of interest, and similar features. Since the indenture covering each issue is an individual and particular contract, some points of difference will usually be found even among bonds of the same general class. Some of the more important classes of bonds are mortgage bonds, debenture bonds, income bonds, equipment trust bonds, and guaranteed and collateral trust bonds.

A *mortgage bond* is issued with certain corporate property pledged as security for the repayment of the loan. In case of default by the corporation

(failure to live up to the provisions of the bond contract), the bondholders have a legal right to take over the property which has been pledged as security. The indenture provides the conditions under which this may be done and outlines the actions to be taken by the trustees on behalf of the bondholders. This recourse, however, is actually of less value than might appear on first thought, because usually if the corporation is not able to earn enough through operation with the pledged facilities to escape default, it is unlikely that the bondholders will fare any better by assuming operation of the property, or that they will be able to obtain a full return from sale of it.

Mortgage bonds may be either a first or second lien against the property. A first mortgage issue has first claim against the pledged property and is therefore a lien prior to all others with respect to the particular property. The interest payments on such bonds also take precedence over all other claims on earnings. Whenever possible a corporation will meet interest payments on its mortgage bonds because default would usually result in litigation, or possibly loss of property, or bankruptcy.

The question may be raised as to why mortgage bonds are issued when such dangers are thereby created. Due to the priority of claims, the regularity of interest payments, and the security behind such issues, creditors are willing to invest in mortgage bonds at a lower interest rate than they would accept from other loans not so protected. From the issuer's point of view, there are many firms for which the general business risk is low (such as public utilities) or whose earnings are relatively little affected by cyclical changes in business. Such a firm may find it cheaper to borrow on mortgage bonds without incurring great risk thereby. On the other hand, there are many types of business to which mortgage bonds are not applicable; e.g., a concern whose capital is largely circulating will have no suitable property to pledge for security. Thus mortgage bonds generally can be issued only on the basis of fixed capital; the ratio of fixed to circulating capital is a limiting factor in determining the size of the issue.

Debenture bonds are issued on the credit standing of the corporation and are backed only by the corporation's promise to repay the loan and to pay interest at stated intervals. The indenture covering such bond issues often carries certain protective covenants which preserve the priority status of the issue as it exists when the bonds are issued or may even call for some improvement in that status. These covenants specify certain actions on the part of the company to ensure its ability to meet interest payments and to repay principal. The debenture does not possess the right of foreclosure and obviously has a lesser claim against the corporation than

mortgage bonds. Such issues, therefore, involve less risk for the company and more for the bondholder and accordingly will usually command a higher rate of interest than mortgage bonds of the same company. However, the debentures of one company may be far superior to the mortgage bonds of another. If a debenture is not preceded by other securities and is properly protected against the subsequent issuance of such securities, it stands as the senior ranking security of that company and may well be superior to first mortgage bonds of another company.

The *income bond* is one of the less common types and one which is not regarded too favorably by buyers of securities, although it is sometimes issued by a company desirous of avoiding excessive fixed annual interest charges. It is a general credit security with interest payments wholly contingent upon earnings. A specific rate of interest, either cumulative or noncumulative, is stipulated in the contract covering an issue of this bond, with provision that the interest be paid at stated intervals provided earnings justify such payment. From the company's standpoint the income bond is an advantageous device for obtaining funds in a business where earnings are likely to fluctuate widely since interest payments do not have to be met in periods of low earnings or losses. However, in order to attract purchasers for such securities a higher rate of interest must be offered than for mortgage bonds, debentures, or other types which yield assured regular returns.

The *equipment trust bond* is a special type of bond widely used, especially among transportation companies. This security is used to finance the purchase of equipment, the customary procedure being for the corporation to make a part payment on the equipment, leaving the remainder to be financed by a trustee who accepts equipment trust bonds in payment from the corporation and offers them to the public. The trustee receives and retains title to the equipment until all bonds are paid off. The usual contract calls for the borrowing company to make periodic payments to be applied toward retirement of the debt. The rate of reduction of the debt through such periodic payments should be more rapid than the rate of depreciation of the equipment; when this is true the bondholders' margin of safety becomes progressively greater as the loan runs its course. The fact that such securities are issued to finance the purchase of equipment for which there is usually a ready market gives the investor an added factor of safety.

Guaranteed bonds are those issued by one company with a second company guaranteeing that provisions of the indenture will be fulfilled. The guarantee adds security to the issue only within the limitations of the specific contract and to the extent that the guarantor is able to fulfill its commitments if called upon to do so. *Collateral trust bonds* are those issued

by one company with securities of another or several companies pledged as protection for the loan. Such bonds may be issued by a holding company which pledges securities of operating companies as collateral. Because bonds of this type are backed both by the pledged securities and by the general credit of the issuer, they are quite likely to be more attractive from an investment standpoint than those which either the holding company or one of the operating companies alone could issue.

Other Forms of Borrowing

Although stocks and bonds comprise the principal means of obtaining original capital, companies may resort to short-term borrowings. After operations have begun, a corporation may procure capital aside from its own earnings from a variety of sources: trade creditors, banks, finance companies, commercial-paper houses, and others. Funds so obtained are used chiefly for circulating capital needs but may also be utilized to some extent for fixed capital needs. Ordinarily such funds obtained on short term cost less in interest than long-term borrowing. Merchandise may be purchased on credit and, in so far as the books show accounts payable covering such purchases, trade creditors furnish a portion of the circulating capital. In a similar manner wages payable represent a contribution to the corporation's capital, since productive labor results in an increase in the value of inventories.

Banks are an important source of short-term funds for most corporations. Through arrangements with one or more banks for a line of credit (maximum amount which bank will loan), corporations may, and frequently do, continuously utilize this method of obtaining capital. By such a flexible method a company may increase or decrease its circulating capital as the needs for such funds fluctuate, thereby ensuring adequate funds for short-term needs (such as financing of peak inventories) and also avoiding a redundance of circulating capital. While it is imperative that adequate circulating capital be available, most corporations guard against a redundancy of funds to avoid unnecessary cost and consequent inefficiency and waste. A corporation needs good banking relations, for banks are frequently the only source of funds in time of crisis.

Finance companies constitute a common means of carrying inventories, although their rate of interest is generally higher than that for bank loans or other direct borrowings. An example of the use of finance companies is furnished by automobile dealers, a large part of whose inventories are financed by loans from such institutions.

Open-market borrowings are an important source of funds for certain companies. These borrowings are made through commercial-paper houses which purchase the company's unsecured promissory notes for resale to large investors or financial institutions. These notes, termed *commercial paper*, are usually issued in large denominations by well-known companies, and therefore this method of financing is applicable only when a large amount of capital is required. The term of such loans is generally about four months and rarely over one year. Because of the thorough investigation made by the commercial-paper house before it purchases the notes, the interest rate on commercial borrowings is relatively low. The chief advantages of this method of financing are wide distribution of the loan, a low interest rate, and good advertising. However, the issuance of commercial paper is largely a fair-weather transaction limited to times of good business. Companies are reluctant therefore to forego their contacts with banks and the establishment of credit lines in order to borrow on commercial paper.

Earnings

During periods of profitable operations a concern may increase its capital by retention of earnings. After all fixed obligations have been paid, a corporation may elect to reinvest a part or all of the remaining earnings in the business rather than pay them to stockholders in the form of dividends. The retention of earnings is a desirable way of providing for the gradually increasing capital needs—both circulating and fixed. The nonpayment of dividends which have been earned does not ordinarily constitute a loss for the stockholders since their equity in the business is increased by the amount of retained earnings.

EQUITY STRUCTURE

The composite of all claims by suppliers of financial resources forms the equity structure of the corporation. It is imperative that this structure provide for adequate capital to be available on terms which the company can afford. A permanent commitment of capital or funds by common stockholders is in all cases necessary as a base on which to build the equity structure. It serves as protection for investors of other types. A no less important reason for the equity represented by common stock is the avoidance of fixed charges, which are embarrassing in periods of low earnings or losses. Common stock represents true venture capital, which risks all for the chance of handsome returns from a prosperous business.

However, total financing by means of common stock is not often either feasible or desirable from the firm's point of view. It would be possible for an enterprise to finance itself wholly in this manner only if abnormally high rates of profit could be anticipated, and even then it would be uneconomical because the capital would be redundant except in periods of peak requirements. On the other hand, additional capital can be realized—at terms more reasonable than the rate necessary to attract venture capital—if the provisions of the contract under which it is to be secured and the basic common stock both offer ample protection to potential investors. By use of bonds, preferred stock, and short-term borrowings, companies may obtain additional funds at a cost below the average earning rate on total capital. In such cases the residual earnings available to common stock are proportionately greater than if the company had been financed by common stock only.[3] Because of the fixed commitments made by the corporation when capital is borrowed, the financial risk of common stock is greater than when such methods of financing are not used. Some companies practice "trading on the equity" (i.e., relying on borrowing for the greatest possible proportion of their capital requirements) as an integral part of their financing technique, but the amount of financial risk the common stockholders are willing to assume limits the extent of this practice.

Equity structure for a business is determined by several factors: (1) the nature of a business, (2) the character of capital needs, including ratio of fixed to circulating, and (3) the financial risks which it can reasonably afford. In general, capital which represents long-term investment by the corporation, whether fixed or circulating, should be obtained on a long-term basis—by stocks, bonds, or other long-term borrowings. Capital required for gradual growth, either fixed or circulating, and for gradual retirement of fixed obligations, may be supplied from earnings. Short-term needs of recurring or fortuitous nature may be financed by bank loans, open-market, or other short-term borrowings.

Going Concern Value

In order to attract the necessary capital for promotion and organization of a business enterprise, the venture must possess economic value, or be able

[3] Also, setting up its equity structure to include bonds reduces the corporation income tax. A capital supplier who invests $2,000 and accepts a stock equity will find that if earnings are 10 per cent before a corporation income tax of 40 per cent, he will get only 6 per cent return, assuming all earnings are paid. If he takes a 10 per cent bond for $1,000 and stock for the other $1,000, he will get an 8 per cent return, assuming all earnings are paid. The difference is due to the fact that bond interest is an expense deductible from earnings in computing income-tax liability.

to produce goods or services capable of satisfying a demand. The amount of capital which will be made available to a new enterprise depends on prospective earnings of the company and the price of capital. Because buyers of securities have alternative opportunities to use their funds, the prospective earnings of a new venture must be sufficient to pay to them the prevailing rate of interest on virtually risk-free securities (e.g., high-grade government bonds) plus a premium commensurate with the risks inherent in the particular enterprise. Otherwise, capital will flow into other channels where such returns are possible.

The capital value of a corporation is an amount determined by capitalizing expected earnings at a rate high enough to compensate for (1) what might be earned on virtually risk-free securities, (2) the risk above this appropriate for the nature of the industry, and (3) the risk above or below that for the industry appropriate for the particular firm. The amount computed in this manner is termed the *going-concern value* of the corporation. To illustrate the procedure of capitalization, suppose that earnings are estimated at $100,000 per year, that the prevailing basic interest rate is 2½ per cent, and that a premium of 7½ per cent is required for the risks inherent in the industry and the firm. The capitalized value of the earnings, or the going-concern value, will then be that sum of which $100,000 is 10 per cent—$1,000,000.

Earnings come into being through the management's use of capital. Management is the vitalizing element which transforms the physical assets of the corporation into a live and productive organization, so that from the complex of fixed and circulating capital will emerge a going concern with earning capacity. Therefore, the capitalized value of a successful corporation's earnings (its going-concern value) should substantially exceed the required investment in fixed and circulating capital.

The going-concern value and the amount of capital required to maintain control of needed instruments of production determine the acceptable upper and lower limits of total capital. However, corporations are frequently found to be over- or undercapitalized. Some concerns are overcapitalized from their inception because of an overestimate of prospective earnings, whereas others become so as a result of reduced earning power or an increase in interest rates. On the other hand, some firms are undercapitalized in relation to prospective earnings through failure to obtain sufficient capital to provide for fixed and circulating needs either at the time of organization or after growth and expansion have occurred. From some points of view, any deficiency of capital below the going-concern value can be interpreted as indicating undercapitalization. The market values of outstanding securi-

ties often indicate the appraisal of buyers as to the redundancy or insufficiency of capital. When an enterprise is overcapitalized the market value of its securities is likely to be below the book value of the equities represented by such securities, and vice versa.[4]

FINANCIAL STATEMENTS

Since the profit motive is the primary aim of business ventures, it is necessary periodically to take stock of the value of the business and to estimate (business) profits which have been earned. The only way in which profits of an enterprise can be computed precisely in terms of money is to compare total returns with capital invested after the business is liquidated. The sales price of the corporation, less the original and subsequent contributions of investors, plus all profits and interest [5] distributed to suppliers of capital, would represent accurately the dollar profits earned during the period of its existence. However, administration of income requires that periodic estimates be made of earnings in order to determine whether the complex of receipts, expenditures, and depreciation of capital assets is resulting in a profit commensurate with the invested capital. An estimate of earnings is also necessary to appraise the company's ability to cover fixed charges and allow some return to stockholders, as well as to determine what distribution of profits can be made in the form of dividends. Finally, a periodic evaluation of invested capital is necessary to indicate the degree of preservation of owners' and creditors' investment.

Income Statement

Accountancy furnishes the means of estimating periodic earnings and measuring values of property and equity accounts. The income statement—commonly called profit-and-loss statement—is a summary of all receipts and expenditures flowing through a business as a result of normal operations within some chosen period, usually one year. The excess (or deficiency) of receipts over expenses chargeable to operations incident to securing such receipts is the operating profit (or loss) for the period.

[4] Book value measures the assets of a corporation, minus liabilities other than its stockholder liabilities.

[5] It should be noted that even this estimate of the profit would contain a mixture of business and pure profit. The excess of value at liquidation over original cost plus dividends, etc., does not measure pure profit because it is uncorrected for price changes. The owners will have made a pure profit if they have secured an amount of money which will command more goods and services than did their investment.

Table 60 COMPARATIVE CONSOLIDATED INCOME ACCOUNT

Royal Typewriter Company, Inc., including domestic subsidiary, years ended July 31

	1944	1943	1940
[a] Net sales	$18,758,594	$13,303,673	$18,781,556
Cost of sales	14,825,676	10,650,205	9,478,163
Gross profit on sales	$ 3,932,919	$ 2,653,468	$ 9,303,393
Revenue from repairs, etc., net	2,068,225	1,071,629	113,169
Gross income	$ 6,001,144	$ 3,725,097	$ 9,416,562
Operating expenses	4,318,283	3,288,113	6,460,961
Operating profit	**$ 1,682,861**	**$ 463,984**	**$ 2,955,601**
Other income	95,720	98,291	43,717
Total income	**$ 1,778,581**	**$ 535,275**	**$ 2,999,318**
Interest, etc.	—	1,907	3,465
Other deductions	—	—	42,553
Federal income tax	722,000	206,928	511,849
Federal excess profits tax	—	—	—
Other taxes	—	—	53,698
Net income	**$ 1,056,581**	**$ 326,439**	**$ 2,387,753**
Preferred dividends	263,844	263,844	263,882
Common dividends	617,821	805,854	1,275,935
Surplus for year	$ 174,916	[b] $ 743,259	$ 847,936
Previous surplus	9,157,113	9,322,628	6,626,374
Sundry credits (net)	37,788	[c] 577,744	—
Reserve for contingencies	—	—	[d] 287,877
Surplus charges	—	—	7,064
Stock dividend (common)	805,854	—	—
	—	—	—
Surplus, July 31	$ 8,563,963	$ 9,157,113	$ 7,179,368

[a] As reported to SEC in years prior to 1944.
[b] Deficit.
[c] Includes $481,634 claim for refund on account of federal excess profits taxes for year ended July 31, 1942.
[d] Provision for possible further losses on investments in and advances to foreign subsidiaries.

General Note: Earnings from war-materials contracts for years ended July 31, 1942 and 1943, were reviewed by War Department and no refund was required. No provision made on account of earnings for year ended July 31, 1944, as company deems none to be required.

Since all items included in the income statement are taken from the books of account in which all transactions are recorded in precise dollar terms, the question may be raised as to why the computed profit (or loss) is not wholly accurate. The value of a company's inventory often consists of both the original cost of some items and present market values of others, and there can be no accurate forecast of what inventory will later sell for; altogether, then, after the income statement is drawn up some income estimates will later prove to be incorrect. Sale of inventory valued at a low cost in a period of rising prices, for instance, will tend to overstate current profits. The inability to determine precisely the portion of certain income and cost items which are applicable to the present period also leads to inaccuracies. Estimates of the provisions that must be made out of current income for future contingencies like bad debt losses are often incorrect. Moreover, the cost of a machine may be depreciated over a ten-year period —10 per cent of the cost charged each year—but, if the machine wears out or becomes obsolete in five years, profits during those five years have been overstated.

Balance Sheet

The balance sheet represents a cross-section view of the property of the corporation and the structure of equities representing claims of owners and creditors in that property; it derives its name from the balance which always exists between the property and equities. It is based upon the fundamental equation: ASSETS (what the firm owns) = LIABILITIES + CAPITAL (what the firm owes to outsiders and to its owners). The balance sheet items show as of the date of its preparation the book value of each property and equity account. In contrast with the income statement, which summarizes the results of operations for a given accounting period, the balance sheet gives an instantaneous view of the entire business in monetary terms. In preparing the balance sheet, adjustments to the property and equity accounts usually must be made in addition to those made in the income statement. The income statement is concerned only with operations whereas the balance sheet valuations depend on nonoperating changes as well. For example, if a company suffers the loss of a building through fire, the income statement would include an entry covering the depreciation of the building chargeable to the given period only, whereas the balance sheet will show the necessary reduction in property account due to writing off of the remainder of the building account and the corresponding reduction in the equity accounts—capital and surplus. As the balance sheet is affected by inaccuracies of the

Table 61 COMPARATIVE CONSOLIDATED BALANCE SHEET

Royal Typewriter Company, Inc., including domestic subsidiary, as of July 31

Assets	1944	1943	1940
Cash	$ 3,621,290	$ 2,253,407	$ 2,580,194
United State securities	[a]2,449,270	[a]1,580,441	—
Drafts and accounts receivable (net)	1,085,317	2,324,623	3,625,727
Inventories	4,771,634	5,098,446	3,816,422
Total current assets	$11,927,511	$11,256,917	$10,022,343
Real estate, machinery and equipment	5,868,828	5,806,777	5,468,448
Depreciation	3,807,082	3,593,525	3,219,533
Net property	2,061,746	2,213,252	2,248,915
Patents, good will, etc.	1	1	1
Foreign investments and advances	10,000	10,000	[b]10,000
Excess profits tax refund claim	481,634	481,634	—
Deferred charges	829,334	597,525	355,785
Total	$15,310,226	$14,559,329	$12,637,043
Liabilities			
Employees' withholding tax, etc.	$ 131,769	$ 131,838	—
Accounts payable	516,481	548,002	$ 220,892
Commissions payable	15,944	24,700	152,701
Accrued wages, etc.	171,718	162,363	244,750
Unredeemed coupons	107,723	81,817	35,681
Federal and state tax reserve	958,955	415,677	765,283
Total current liabilities	$ 1,902,590	$ 1,364,398	$ 1,419,307
Contingency reserve, etc.	—	—	—
7% preferred stock ($100)	3,769,200	3,769,200	3,769,750
[c] Common stock	1,074,472	268,618	268,618
Surplus	8,563,963	9,157,113	7,179,368
Total	$15,310,226	$14,559,329	$12,637,043
Net current assets	$10,024,921	$ 9,892,519	$ 8,603,036

[a] Includes United States tax notes due Sept. 1, 1945: 1944, $447,568: 1943, $603,720. Accounts certified by H. E. Bacon & Co. (formerly A. H. Whan & Co.) in 1944, and by A. H. Whan & Co. in prior years.

[b] After $287,877 reserve (see income account) provided in 1940 fiscal year.

[c] Represented by 1,074,472 $1 par shares in 1944, and $268,618 no par shares in prior years.

income statement and other nonoperating (capital loss or profit) adjustments, it must be recognized that the valuations contained in it are simply estimates.

In estimating the worth of a business, various concepts with their concomitant principles of valuation may be used. Actual cost, based on prices paid—adjusted for actual depreciation depletion, or loss of property, plus the excess of receipts over expenditures—yields an *historical cost value* for an enterprise. Valuation on the basis of *replacement cost* (based on prices currently prevailing) would yield a different set of property and equity values. The ordinary balance sheet is usually a compromise between these two types of valuations and must be interpreted accordingly. A third method of valuation is that of determining going-concern value, and this method may yield a total value far different from that obtained by using either historical cost or replacement cost.

Comparative Statements

Careful analysis and interpretation of both the balance sheet and income statement are necessary for proper administration of corporate activities and income and for wise purchases of corporate securities. Management must know whether the proper balance is being maintained among asset and equity accounts, whether sufficient working capital is available to protect the credit standing of the company, whether earnings are commensurate with capital invested, and whether there are acceptable reasons for changes which have occurred in the accounts. Statements for the current period alone are insufficient for a complete understanding of the financial conditions of a business, and therefore the usual practice is to construct comparative statements covering the last few years. The analysis of such statements indicates growth or decline of a company as well as changes in the various balance-sheet and income accounts. A useful adjunct to the comparative statement is the *percentage statement* which shows the proportionate distribution of balance-sheet and income totals among the various accounts, a device which improves comparability between years. The *percentage balance sheet* shows the distribution of capital among the various asset and equity accounts, whereas the *percentage income statement* indicates the relative effects of various income and expense items on the realized earnings of the period.

FINANCIAL RATIOS

In addition to the percentage statement, there are certain balance-sheet and income ratios which are of considerable assistance in appraising the current financial condition of a company. Questions regarding the trend of the working capital position, length of the circulating capital cycle, stockholders' equity, investment in plant and equipment, and earning power—all may be answered by ratios computed from balance-sheet and income accounts. For example, the short-term credit rating is often evaluated by the "working capital" and "acid test" ratios. The *working capital ratio*—circulating capital divided by current liabilities—is a measure of the adequacy of circulating capital, whereas the *acid test*—ratio of cash and receivables to current liabilities—is a measure of the firm's ability to liquidate its current obligations. Such ratios, however, are not infallible criteria, for earning power is the ultimate test of a company's ability to meet its obligations. Space does not permit a detailed examination of the use and limitations of the various financial ratios which are treated fully in standard accounting works. It must always be borne in mind that ratios express financial relationships which are aids to business judgment—they are only supplemental to financial statements.

BUSINESS FAILURE AND REORGANIZATION

Many firms suffer reverses or failure and thus present special problems in corporation finance. Some have never enjoyed a chance to succeed, being doomed to failure from their inception because of inadequate or improper financing or because no economic justification for the venture exists. Some firms fail as a result of management's incompetence or malfeasance. Corporate failure is not an event that occurs overnight, but results rather from progressive deterioration of a company's earning power and financial condition. The causes of failure are numerous—inability to operate as efficiently as competitors, changes in demand for product, overexpansion, adverse general business conditions, fortuitous disasters, financial weaknesses, and defects of management. The effects of such causes often go unnoticed for a long time until the corporation actually defaults on some financial obligation. Default is, however, evidence of the fact that the company has failed rather than the cause of failure. If competent management recognizes symptoms of impending failure and takes steps to

strengthen the financial structure or improve operations and markets, actual failure may be averted.

Corporate failure may be either economic or legal in character. When earning power is insufficient to return the prevailing interest rate on capital plus the appropriate premium for risk, a business is an economic failure. On the other hand, legal failure occurs only when a corporation is unable to meet its financial obligations.

A company financed without bonds may be an economic failure and yet continue in business indefinitely without becoming a legal failure. If the management has firm control, it may, merely to maintain jobs, continue the business even though it does not cover much more than average variable cost. It may not cover the annual depreciation or allow anything for dividends; it may, in fact, spend reserves and reduce working capital in order to continue operation. In such cases, the interest of management may conflict with that of stockholders, whose equity is dissipated by continued operation at a loss. The stockholders' recovery would be greater if the firm were liquidated before it reached a condition of legal failure.

Legal failure may be of either a temporary or permanent nature. Frequently firms which are technically insolvent in periods of depressed business conditions again become solvent with returning prosperity, provided they have been able to escape actual default. One device sometimes used in this situation is an arrangement for a moratorium—period of delay—on some of the obligations. However, where a corporation is forced to default on its obligations more drastic action is required. It may be necessary to enter bankruptcy, with either subsequent reorganization of the financial structure or dissolution of the corporation.

Before 1933, corporate failures usually resulted in considerable loss to all investors save those with legally enforceable contractual claims, because there was no legal procedure to authorize a reorganization equitable for all parties concerned—the corporation, creditors, and stockholders. Often, settlement could be made only through foreclosure and sale of the properties because dissenting creditors or stockholders were able to prevent a reorganization approved by the majority. This meant the dissolution of many companies when it was unwarranted and unjustifiable. To correct this situation, the Federal Bankruptcy Law was amended in 1933 and 1934 with the enactment of section 77 and section 77B, intended (1) to assure proper distribution of proceeds from the sale of assets among all creditors so that none would obtain an unfair advantage, and (2) to relieve unfortunate but honest debtors from further obligations. Section 77 specified procedures for the reorganization of railroads, and its provisions were extended

to cover business corporations by section 77B. In 1938, on the basis of four years' experience, an improved procedure for reorganization of business corporations other than railroads was established under Chapter X of the Bankruptcy Act.

The procedure as outlined by the present law permits the reorganization of many companies which formerly would have had to suffer the stigma of bankruptcy and the travail of foreclosure. The law requires the appointment of a trustee after a firm has been adjudged bankrupt by a federal court if the liquidated and noncontingent liabilities are above $250,000. The trustee is required to prepare a plan of reorganization of the financial structure, or alternatively, to demonstrate that successful reorganization would be impossible. The plan must set forth the properties of the corporation, the allowed claims against the corporation, a proposed new equity structure, and the treatment to be accorded all creditors and stockholders—both assenting and dissenting. The plan is submitted to the court for approval and in certain cases must also be referred to the SEC. It is then submitted to the various classes of creditors whose claims have been allowed, and in the case of solvent corporations, to stockholders as well.

Before any plan can be placed in effect it must be acceptable to creditors representing two thirds of the amount of each class of indebtedness and to a majority of each class of stockholders affected. In such plans of reorganization, some alteration or modification of the rights of creditors generally or of some class of them is mandatory. There is usually a scaling down of fixed charges and issuance of securities with lower priority or less certainty of income than those held by the old creditors or stockholders. For example, debenture bonds may be issued to replace mortgage bonds, preferred stock to replace debenture bonds, common stock to replace preferred, etc. Although it may even be necessary to exclude the old common stockholders completely, it is generally conceded that they should have some rights in the new company and be allowed participation on some basis. Unsecured creditors (those without a claim against any specific property) are also provided for under this procedure, usually by receiving bonds or stock.

When there are only unsecured claims against the corporation a similar if somewhat less formal reorganization is provided for in Chapter XI of the Bankruptcy Act. In all cases, however, reorganization is also subject to the state laws as well as to the Federal Bankruptcy Act.

THE MARKETING OF CORPORATE SECURITIES

The market for securities of a newly organized firm is generally somewhat restricted because its earning potential is unknown and its standing in the business community has not been established. New securities of an established concern may be offered to old security holders, customers, employees, private institutions, personal friends of the management, or to the general public, whereas the market for securities of the newly organized enterprise is often limited to friends of the management and the general public. In new stock issues for a going firm, old stockholders are allowed to buy in proportion to their present holdings so that they may preserve their equity status and relative control and so exercise the fundamental rights of stock ownership. Securities are usually offered to customers and employees in order to create good will rather than to obtain capital.

Securities may be offered to prospective buyers by direct sale or through an agent, known as an "underwriter," which may be a single investment banking house or a syndicate formed by several investment houses. Offerings to the general public are customarily made through an underwriter. A contract is negotiated between the corporation and the underwriter, describing the securities and specifications of the sale of the securities, including the compensation to be paid the underwriter. The corporation sells the issue to the underwriter at a figure somewhat below the specified offering price to the public, and the difference between these two figures constitutes the agent's commission. In some instances corporations find it to their advantage to sell an entire issue to a single institution, such as an insurance company, rather than to make a public offering. This type of private selling, possessing the advantages of speed and economy, has been especially popular since the passage of federal legislation governing the sale of securities.

FEDERAL CONTROL OF SECURITIES

Before 1933 there was no adequate control over the issuance of securities; consequently their sale was sometimes accompanied by widespread fraud and misrepresentation. Most of the states had ineffectual "blue-sky laws" which attempted to prevent the sale of flagrantly unsound securities representing little more than so much "blue sky." State laws also required the filing of certain information on the financial condition of the issuing company and in many cases required official approval of securities before

they could be sold. Not only was control of suspect interstate offerings generally ineffectual, but also conflicting state requirements hindered the marketing of sound issues. Securities were often offered without sufficient explanation of the risks involved, or their worth was misrepresented by false information. In many instances the advertising was merely misleading, but on occasion actual fraud was perpetrated on hapless buyers. These conditions in the marketing of securities became progressively worse and were a contributing factor to the stock-market crash of 1929. The crash brought to light many of the shady practices of some promoters and investment banks and stirred the ire of Congress. The result was the enactment of a series of federal laws designed to protect the investor from being victimized by unscrupulous operators.

Securities Act of 1933

The Securities Act of 1933, the first of these laws, was enacted "(1) to provide full and fair disclosure to prospective investors of the character of new securities or new offerings, and (2) to prevent fraud or misrepresentation in the sale of securities, old and new." The intent of the law was not to prevent speculation in securities but rather to make sufficient information available to let the prospective buyer know whether he was speculating or investing in a sound security. For this reason, the law is often referred to as the "Truth in Securities Act." Authority for the legislation derives from the rights of Congress, as given in the Constitution, to regulate interstate commerce and control the mails.

To accomplish the first of its stated objectives, the act requires the filing of a registration statement with the SEC describing details deemed necessary to enable the buying public to form an accurate judgment of the value of the securities offered. Furthermore, the act requires that a prospectus, containing information similar to that in the registration statement, be made available when securities are offered to the public. Civil liabilities and criminal penalties are imposed for the sale of any securities by means of fraud or misrepresentation. The outstanding feature of the act is its provision for compulsory registration of all securities (except as specifically exempted [6]) with the SEC if mail or other interstate commerce channels

[6] A partial listing of the securities exempted by the act from registration includes those sold or publicly offered prior to its passage, government securities, securities of interstate railroads under the jurisdiction of the Interstate Commerce Commission, securities issued by receivers or trustees in bankruptcy, certain securities exchanged with old security holders, issues sold wholly to residents of a single state where the issuer does business wholly within that state, and issues amounting to not over $300,000.

are used to sell or to offer to buy these securities or to carry the securities for sale or delivery after sale. Unless the Commission issues a stop order, the registration becomes effective at the end of a brief waiting period and the company can then offer its securities for sale to the public. Responsibility for registration presumably lies with the issuer, although any seller may be liable to a purchaser.

After registered corporate securities have been initially sold to the public, either directly or through an underwriter, they may be freely traded. Trading may be carried on by over-the-counter transactions or through an organized secondary market. The outstanding example of secondary markets is the New York Stock Exchange, which lists more than two thousand different stock and bond issues and which often handles transactions in excess of a million shares per day.

A securities exchange is a market place where buyers and sellers, through authorized representatives (brokers), meet to buy and sell securities. A security must meet the standards of the exchange before it can be listed, and the facilities of the exchange can be used by traders in the security. An exchange will accept a security for listing only after the issuer has filed detailed information concerning the management, financial condition, and operations of the company, although acceptance does not presume the exchange's approval of the security as an income-earning asset. Furthermore, the issuer must agree to conform to the rules of the exchange and to make public periodic reports on the financial status of the company. Listing is limited to the securities of larger corporations having an active market.

Transactions involving the many securities not listed on any exchange must be made over-the-counter, by personal contact between buyer and seller or through a brokerage house. Many corporations seek to list their securities because it increases marketability and therefore enhances the value of a security. The sale of new issues is aided by listing even though the initial sale cannot be made through an exchange.

Securities Exchange Act of 1934

Before 1934, trading in securities was subject to many abuses and undesirable practices by speculators; the prices of individual securities and even the level of the whole market were often forced up or down for the sake of trading profits. These profits were sometimes made at the expense of small and uninformed buyers of securities. False impressions of activity were created by such devices as "wash sales" or "matched orders," in which no change in ownership occurred because the seller of a security matched

his sales with purchases. Speculators used such devices to create interest in a particular security and would sell when the ensuing demand forced the price up. By creating uneasiness through disseminating false reports, speculators were able to force the price of a security down and then to purchase it at a low price. The formation of buying or selling pools was a common characteristic of the market. Through pooling of a large number of shares and using a regulated selling program, the price could often be forced down, or conversely—by pooling purchase orders—forced up. Insiders within a corporation were often able to buy or sell on the basis of advance information concerning corporate actions likely to affect prices, such as impending increases or passing of dividends. Of course, even without such practices it is the nature of trading that as a result of speculation the market prices of securities are frequently far out of line with their economic value.

The malpractices of traders and the ineffectiveness of control by the exchanges—and especially the harmful speculation of the late 1920's before the crash—led to the passage of the Securities Exchange Act in 1934. The purposes of the Securities Exchange Act are to (1) establish federal regulation over securities exchanges and markets, (2) prevent unfair practices on such exchanges and markets, (3) discourage and prevent the use of credit in financing excessive speculation in securities, (4) compel corporations to furnish adequate information as to securities publicly traded in, and (5) control unfair use of information by corporate insiders.

The Securities and Exchange Commission was created to administer the act and was empowered to regulate securities exchanges and markets and to prescribe the conduct of corporations in all matters pertaining to their securities. The act prohibits certain specified practices and types of transactions, requires the supplying of detailed information, and establishes control over manipulative practices and devices, e.g., short sales.

Since excessive speculation is often financed largely by bank loans, the act empowers the Board of Governors of the Federal Reserve System to establish and alter margin requirements—the amount of credit allowed in purchasing securities listed on a national exchange—and to control the lending of money on securities by member banks; it also imposes limitations on borrowing by exchange members, brokers, and dealers. To minimize the advantage accruing to large owners, directors, and officers, certain restrictions are placed on buying and selling of a firm's securities by its insiders. The Commission is also authorized to establish rules and regulations governing the solicitation of proxies.

Other Restrictive Legislation

With the passage of the Trust Indenture Act of 1939, the SEC was granted authority over contracts under which bonds are issued. The act specifies what information must be included in an indenture, establishes eligibility standards for the trustee, and outlines his duties and responsibilities. Registration cannot become effective for securities covered by the act unless the indenture meets the prescribed standards.

Effect of Federal Legislation

Federal control of corporate financing and the marketing of securities has on the whole achieved salutary results in business finance. The provisions of law which require full disclosure of corporate affairs and financial condition have led to improved accounting and the issuance of more intelligible corporate financial statements. This has been valuable to both the investing public and the corporation by focusing attention on the soundness of the company's financial structure. In some cases the preparation of required reports has disclosed financial weaknesses which otherwise might have gone unnoticed and consequently has led to corrections.

Since the various acts were designed not to eliminate speculation but rather to control it, they have abolished many evils in trading and have made the exchanges more nearly fulfill their primary function as a free market place for buyers and sellers. Powers of control vested in the SEC and in the Federal Reserve authorities operate to curb violent fluctuations in market prices, although price trends based on economic factors will prevail in spite of regulation. Furthermore, although security buyers are protected against fraud and presumably given sufficient information to appraise the soundness of securities, the law does not actually protect the buyer from purchasing unsound securities since no guarantee is given concerning their investment qualities. The law simply requires that a security be described without misrepresentation, and by this provision some corporations are discouraged from any attempt to sell worthless or grossly overpriced securities. In general, however, buyers feel greater confidence in new securities offered for sale, and this in turn simplifies the issuer's problem of marketing them. The fact that a security can meet the requirements for registration and listing on an exchange tends to enhance its marketability and render it more acceptable to buyers.

Despite its salutary effects federal control over corporate financing and

securities markets is not without its disadvantages. Much criticism has been directed against the delays caused by the waiting period and the costliness of registration; to avoid this many companies have been driven to private placement of their securities. In the case of small issues the cost of meeting the requirements of the law is proportionately greater than for large ones. Originally all issues over $100,000 had to be registered, but experience proved this to be too stringent, so the exemption limit was recently raised to $300,000. The waiting period is a twilight zone of enforced inactivity during which certain information regarding new issues can be circulated but no offers to buy or sell can be made. Confusion is sometimes created by the differences between federal and state laws and by the conflict in jurisdiction among such regulatory agencies as the SEC, Interstate Commerce Commission, Federal Power Commission, and state utility commissions.

CHAPTER 27

Concentration and Regulation of Industry

In recent times, and especially since the 1870's, considerable attention has been centered on economic concentration—the aggregation of economic resources by the largest firms. This attention is attested by numerous laws now on the statute books, to be described below, and by a continuing program conducted by some branches of the Federal government to encourage small business or to break up corporate giants.

A thorough investigation of economic concentration was conducted by a special committee of Congress from 1938 to 1941, called the Temporary National Economic Committee (TNEC). Its economic studies indicate that there exists a high degree of concentration according to the usual standards of measurement: (1) number of firms in an industry, (2) control of assets and income, or (3) dominant position of one or a few firms with respect to output and employment.[1]

[1] Temporary National Economic Committee, *Competition and Monopoly in American Industry*, Washington, 1940.

Eighty-seven per cent of the corporate assets of all corporations reporting to the Bureau of Internal Revenue in 1935 were owned by less than 5 per cent of the companies filing such returns, and less than 4 per cent of them received 84 per cent of all corporate net income. Of all the corporations reporting from every part of the country, $\frac{1}{10}$ of 1 per cent of them owned 52 per cent of the total assets and earned 50 per cent of the total net income. A report of the National Resources Committee indicates that the 200 largest nonfinancial corporations owned 57 per cent of the assets of all such firms in 1939, as compared to 49.4 per cent in 1929. These data show that ownership of corporate property and control of corporate income are now concentrated in the hands of a relatively small number of giant corporations. However, some reduction in the extent of concentration in manufacturing is evident when 1946 is compared with 1939.[2]

In output, the 200 largest manufacturing concerns produced in 1935 over one third, by value, of the total output of all manufacturing concerns. Of the 275 industries listed by the *Census of Manufactures*, there were 29 industries in which the 4 largest firms accounted for over 75 per cent of the total output, 45 in which the 4 largest firms accounted for over 66 per cent, and 87 in which the 4 largest accounted for over 50 per cent of the total. In the same list of 275, there were 4 large industries, 10 medium-sized and 117 small ones, where the 8 largest firms employed more than 50 per cent of all workers engaged in the industry.

Since one particular item is frequently produced by only a few of the members of an industry, the extent of control over an individual commodity may often be greater than that indicated by industry-wide studies. Data on 1,807 different products taken from the *Census of Manufactures* for 1937 indicated that over 85 per cent of the total output was concentrated in the hands of the 4 largest firms, who were furthermore engaged in the manufacture of over 25 per cent of the total number of the products. In the case of 291 of the listed products, a single leading producer accounted for from 50 to 75 per cent of the total output.

In the field of marketing, concentration exists to a lesser degree than in manufacturing. Nevertheless, there has been a significant trend toward large-scale merchandising in recent years, as evidenced by the fact that the list of the 250 largest corporations in the United States in 1935 included 9 trading companies. The concentration of power in this field has not advanced far, however, since independent merchants, each with a single

[2] See *Survey of Current Business*, United States Department of Commerce, November, 1947.

store, owned nine tenths of the retail outlets and made two thirds of the sales in 1935.

Some reasons for interest in—and usually hostility to—concentration are the belief that large firms restrict competition to the disadvantage of buyers and competitors, and the fear that the economic power of large concerns will be utilized to influence government. American economic history affords enough cases to justify these fears in some measure. Opposition also stems from the tradition of initiative, enterprise, and independence of the American people rooted in the history of the frontier. Reaction against concentration is therefore partly the expression of a protest against the change from an expanding agricultural civilization to modern industrialism.

The problem of concentration is a perplexing one, for bigness in business—as in labor unions—produces mixed results. The large-scale producer is frequently more efficient in some lines of production than the small-scale producer. Large firms usually pay higher wages and salaries and provide better working conditions and security benefits for employees than their smaller competitors.

Despite these conceded advantages for workers, investors, and the total economy, big business is the object of attack from all sides. Labor leaders are fond of stressing the evils of monopoly in business, and justify monopolistic practices of unions on the ground that they are necessary to combat the power of big business. Farmers traditionally have opposed the financial power of large industries. The Populist party of the late nineteenth century attacked "Wall Street" as the consummate evil in the American economy. Small business in general has shared this view, looking upon big business as a threat to its continued existence. This actual or latent hostility finds expression in the political trends of the twentieth century, and was brought to a climax in the restrictve legislation of the New Deal administration.

The growth of big business raises a fundamental issue relating to the operation of the national economy. Among them, two or three hundred corporations wield enormous economic power. When power is great, the many who do not share in its control must trust the wisdom and good will of the few who do. In contrast to this situation, the economy has traditionally been viewed as competitive, composed throughout of small producers in competition with each other, each exerting only a small influence on production, price, and income. Market influences—price-cost-profit relationships—keep small producers in line. This process can be regarded as self-regulative and more or less automatic.

The danger that concentration of economic power will lead to exploitation of the public, in terms of the prices charged for goods or the prices

paid to suppliers of materials and productive services, poses the problem of controlling monopoly and monopolistic practices. It is seldom clear, however, whether gigantic size is an unmixed evil. It was recognized in some of the earliest attempts to regulate monopoly that the existence of monopoly in some fields is unavoidable. Thus, the railroads, grain elevators, and similar businesses were recognized as "natural" monopolies and were placed under control to assure their operation in the public interest. When the Sherman Act was passed in 1890 to outlaw monopolistic agreements, regulation became concerned with breaking up the combinations which caused concentration. Later legislation restricted some of the practices most commonly used by business to build up monopoly power. Some New Deal legislation resulted in government-owned electric power production and distribution facilities with the ostensible design of providing a "yardstick" of government competition with which to measure the performance of already regulated monopolies. In the case of labor unions, however, governmental policy was directed towards encouraging the degree of monopoly power inherent in union organization, and the right of labor to organize has been accorded the protection of the Federal government since 1933.

Despite the many forms it has taken, however, government efforts to prevent business monopolies have generally been directed toward changing the conditions under which the businesses operate. Outside the field of public-utility regulation,[3] the control exercised has not prescribed prices, which constitute the heart of business contracts. Control of monopoly has instead been directed at the conditions under which contracts may be made, thus permitting a wide variety of alternative choices to be made by the firms subject to regulation. Regulation of competitive industries like agriculture, on the other hand, has been motivated by the desire to raise prices for specific commodities and has been directed at the price itself. Public attitude toward the proper sphere of government action helps to explain the difference of approach. Government control of prices was unthinkable when industrial monopoly became the target of legislation in the latter part of the nineteenth and the first part of the twentieth centuries. But by 1930 many of the inhibitions against such control had been removed, especially when farm prices came under legislative influences.

ECONOMIES OF SIZE

The American antimonopoly policy is still rooted in its nineteenth-century origins, and could be expected to change as new techniques and

[3] See Chap. 30.

new attitudes are brought to bear upon it. A major change in attitude has to do with the question whether bigness itself necessarily results in an uneconomic use of resources and in abnormally high prices for the products produced by the largest concerns.

True Economies vs. Monopoly Profits

Mass production, as a result of specialization of labor, employment of machines, and spreading of overhead expense, results in lowered unit costs of production. The reduced costs permit lower selling prices which may be, and usually are, reflected in an enlarged volume of sales, as might be expected from the usual characteristics of demand schedules. In this manner the economies of large-scale production, where they occur, contribute to a higher standard of living. There is, however, an optimum production level beyond which average, as well as marginal, costs of production tend to increase as larger plants are built; expansion of size beyond that point becomes less economical.

The profit motive is the impelling reason for the expansion of individual firms, just as it is for all business activity. Expansion up to a point provides the means of earning greater profits through the economies of larger scale operation. The existence of high profits, however, is not necessarily proof that a firm is economical or efficient in operation since it may enjoy monopoly power. In general, maximum monopoly profits are realized on an output smaller than that which would yield the lowest cost of production. Under such conditions the full advantages of large-scale operations are denied to the economy as a whole. If the products in question could be produced under conditions of pure competition without sacrifice of technological advantages, the public would benefit by the removal of monopolistic influences. In many cases, however, this is not possible; several big units are often more efficient than a large number of small units.

Economies of Large-Scale Operations

Large firms, whether competitive or monopolistic, enjoy certain advantages denied smaller ones. It is only the large firm which is able to take full advantage of the economies of specialization and division of labor. High-volume production permits long-range planning and permits a degree of diversification sufficient to provide continuous operation, eliminating to some degree the alternate periods of activity and idleness which so often characterize the small specialized business. Mass production methods, as

exemplified by assembly-line techniques, permit the breakdown of production work into routine tasks and assignment of these tasks to workmen on a full-time basis. This enables the laborer to become skilled at his particular task and contributes to the increased productivity of labor, thus reducing costs of production. Continuity of operation through diversification and large-scale output make possible the use of specialized machinery, thereby lowering the unit cost of production.

The large firm, through its ability to purchase in large quantities and to space its orders so that they fit in advantageously with the suppliers' own production schedules, can effect substantial economies in its purchase of materials and supplies. Utilization of by-products is a further economy often denied the small firm. In some industries, notably the meat-packing industry, the margin between cost and selling price is so slight that full utilization of by-products is ordinarily necessary in order to show a profit on operations. Furthermore, the large firm is usually better able to pay the salaries required to attract outstanding men who possess high managerial ability; and it can also afford costly experimentation and research which may lead to improvement of product, discovery of more advantageous marketing methods, and lowering of production costs. While it would appear to be physically possible for a combination of small firms to achieve, through co-operative efforts, many of the advantages of large firms, as a practical matter such co-operation is seldom realized. A major obstacle to such co-operation is the individualistic nature of the entrepreneur in most small firms.

Large-scale operation, however, is not without its disadvantages, particularly its comparative inflexibility. Because of its size the large firm necessitates the delegation of authority as well as establishment of routine operations. The delegation of authority entails rules and regulations that often prove cumbersome when contingencies arise. Moreover, routines set up in production or marketing are likely to be retained even after they become manifestly inefficient, because a change on a large scale is regarded as costly or difficult. Mere size may engender inertia unless there is active competition in an industry. In its financial structure, the heavy investment of the large firm in fixed assets imposes fixed charges that may prove a hardship during periods of depression or whenever production is slackened.

Economies of Combination

The combination of several firms for the purpose of effecting economies or gaining mutual advantages in production and marketing is a common

means of obtaining large-scale business operation. Combination may be effected through the merging of firms at the same stage of industry—*horizontal combination*—or the uniting of different stages of production—*vertical combination.* A combination embracing several automobile assembly plants (General Motors) or grocery product manufacturing plants (General Foods) is of the horizontal type; one including iron and coal mines, smelting plants, and rolling mills (United States Steel) is of the vertical type.

A horizontal combination results in the elimination of certain costs and a reduction in some others. If the combined firms are located in widely separated areas, merchandise can be delivered from several points of production, thus reducing transportation costs. Consolidation of selling, purchasing, engineering, and research staffs effects a further reduction in expenses both directly and in more efficient operation. Another economy, important in periods of low or declining output, can be achieved by operating the more efficient or strategically located units at or near their optimum capacity and closing the other plants.

Vertical combination co-ordinates manufacturing processes at various levels. This co-ordination is reflected in lowered production costs and steadier employment throughout the industry. By control of firms at a lower stage of production, a continuous supply of materials is provided to meet the requirements of each successive stage, thereby necessitating only a minimum inventory. The integration of firms eliminates the need for selling and advertising expense at intermediate stages, it also avoids competition in the purchase of raw materials and component parts, thus effecting worth-while savings in so far as it permits more stable production. (Too often, however, it results in monopsonistic buying.) When the "end firm" is engaged in the production of a commodity for which the demand is relatively constant, integration is advantageous to the firms in the lower stages of production because it gives them an assured market.

The full advantages of large-scale production, however, cannot always be attained through combination. The economies of horizontal combination are limited by the extent of product differentiation among the previously separate firms. For example, when a brand has been previously well established it may be difficult, even unwise, to discontinue its production. In the case of a vertical combination, firms at the lower stages of production are dependent on successful marketing of the commodity produced by the firm at the highest stage. When sales of the final product fluctuate widely the effects are felt throughout the entire organization because the firms at lower stages have no other outlet. This is one of the principal disadvantages of integration. During depressions, an integrated organization may

not be able to take full advantage of prevailing low market prices on materials because it must hold together its own organization for supplying raw materials, even if this means paying prices above the market. It is almost impossible to arrange production schedules for all firms within the combination so that either excess capacity or insufficient production will not be encountered at some stage. The optimum rate of production may be higher or lower for firms at different stages than the rate necessary to assure the optimum rate of production for the finished product.

Economies of Monopoly

The economies so frequently attributed to monopolies are merely those of large-scale business. Most of the outstanding monopolies are large-scale enterprises, although great size is not requisite to monopoly control. Monopoly is dependent on the extent of control of the market rather than on absolute size of firm. The concessionaire who has the exclusive right to vend refreshments in a park enjoys a monopoly just as surely as does the company furnishing transportation service to a city.

Certain economies might be available to the monopolist as a result of his control of the market. He is able to avoid duplication of production and marketing facilities. Monopoly also permits greater standardization of product and elimination of competitive advertising costs, although of course advertising designed to increase consumption may be continued. Under monopoly conditions the costs of a product might be reduced below the level which would prevail under competition, and if a decrease in selling prices ensued, monopoly would then contribute to consumer welfare. These possible economies, however, are more than likely to be outweighed by the lack of competitive pressure to lower costs. The competitive producer is constantly compelled to lower his costs; the monopolist may become sluggish, and produce at higher costs, through lack of competitive compulsion.

From the public viewpoint, monopoly is undesirable. Through his control of the market the monopolist may sell at a price higher than that which would have been established under competitive conditions. Output is held by monopolists to those amounts which can be sold at predetermined prices calculated to maximize profits. These conditions tend to result in inefficient operations and less than full utilization of economic resources. As monopoly spreads throughout industry, the field of free enterprise diminishes, business activity becomes less self-regulatory, and governmental regulation becomes necessary.

TECHNIQUES FOR CONCENTRATING CONTROL

Concentration of industry may be achieved by the expansion of a single firm, the combination of competing ones, or by means of either formal or informal agreements. In an expanding industry one firm may become dominant by increasing its share of the total market without any change in the number or ownership of competing firms. When control is achieved in this manner, and without the use of unfair trade practices, it does not appear to be in violation of the antitrust acts.

Ownerships

The holding company constitutes a means of obtaining control over a vast quantity of corporate assets by a relatively small investment in common stock of the parent company. In some cases the holding company purchases only enough common stock of operating companies to ensure control, whereas in other cases complete ownership is effected by purchase of stock or by exchange of holding-company securities for those of the operating companies. Those who control a firm with a small total investment in the holding company may retain control by issuing nonvoting securities in exchange to stockholders of the operating companies.

The holding-company device has been widely used in the public-utility field and has resulted in the formation of large organizations with subsidiaries operating throughout the country. This method of combination has also been used in the industrial field to form some of the largest corporations, such as General Motors Corporation and United States Steel Corporation. Holding-company combinations affect competition variously, depending on the extent to which operations are consolidated, the extent to which uniform policies are established for all subsidiaries, and the proportion of total output represented by consolidated companies. The legality of such combinations depends on the degree to which competition is lessened among the member firms as well as the extent of control of the industry as a whole achieved by the combination.

In the trusteeship (a device used in the 1870's and 1880's), authority was delegated by the stockholders of various firms to a board of trustees, who in turn controlled the policies and operations of all the member firms. It was this device, first used by the Rockefeller interests in the oil industry, which gave the name "trust" to large monopolistic business combinations. Such a scheme was effective only when there was a high degree of uni-

formity among the various firms with respect to policies and operations. This arrangement, however, violates both the antitrust laws and also the statutes governing corporations, since corporation charters do not permit redelegation of authority originally granted to stockholders.

A single firm may gain control of an industry by outright purchase of the assets or stock of other firms without the formation of a separate holding company. Outright purchase results in a complete merger of all firms and consequently absolute control of operations. In some instances predatory tactics have been employed to force competitors to sell, as in the formation of the original Standard Oil Company (which was later forced by court action to split up into several companies).

Ownership of patent rights on basic machinery or production techniques, or ownership of trade-marks, may give the owner so great an advantage that other firms are unable to compete. Patents are of course sometimes licensed to other firms. In granting a license to use its patents, the owner may partly govern the activities of licensees. However, tying contracts which tend to lessen competition—such as agreements prohibiting a lessee from using or dealing in commodities of a competitor of the lessor—are illegal under the Clayton Act of 1914. After the expiration of an original patent, control may be continued through patents on improvements to machines or processes without which competing firms could not profitably continue production. In a similar manner the establishment of a strong consumer preference for branded merchandise bestows a high degree of control on the owner of a brand name. Thus one of the effects of the federal patent law, which was designed to encourage invention by rewarding ingenuity, is to place the government in the anomalous position of making monopolization easier.

Trade Associations

The trade association is an organization created exclusively for service to the members of an industry. Since membership is wholly voluntary, the association is likely to be strong only when its measures or policies are popular with members of an industry. Many of the activities of the trade associations are legitimate under present laws and tend to improve competition and benefit the public. Promotion of research, aid in settlement of labor disputes, representation of the industry in public negotiations, collection and dissemination of statistical information, and recommendations to improve operating procedures—all are within the law and of value to the

industry. Other activities, such as publication of prices in advance of the time when commodities are sold, and the establishment and administration of marketing and production quotas, are considered as violations of antitrust laws, although even in their legal activities trade associations may cause uniformity of operations and pricing among their members.

The pronouncement of the "rule of reason" by the Supreme Court, in 1911, indicating that interpretation of the laws did not prevent all forms of co-operation between competing companies, gave rise to a rapid increase in the formation of trade associations. During World War I, a further impetus to trade associations was engendered by the War Industries Board in its appeal to associations for assistance in maximizing the production of necessary war goods. The conditions resulting in the National Industrial Recovery Act and the production programs of World War II offered an even greater opportunity for trade-association growth. The need for close contact between government and industry demanded spokesmen for industry; this function was delegated to the associations as a general practice, particularly with respect to problems of production and price control.

It was, of course, during the life of the NRA that trade-association activity skyrocketed to an all-time high level. Under the NRA, formulation and administration of industry codes were entrusted to trade associations in those industries where competent associations existed. Administration of codes encompassed price control, allocation of markets, and establishment of production quotas. In many cases there was a restriction of output for the purpose of obtaining higher prices, deemed necessary for recovery from the effects of the depression. To a large extent antitrust enforcement was suspended during this period.

Informal Arrangements

Concentration of power may also be achieved through less formal devices, such as interlocking directorates or price leadership. When a few individuals become directors of several companies operating within the same industry, their control may be practically as complete as if a holding company or trust form of combination were in effect. Interlocking directorates which tend to lessen competition violate the Clayton Act.

When one company becomes outstanding in the manufacture of a product, all other companies may voluntarily elect to follow its leadership. A conspicuous example is the price leadership enjoyed by the United States Steel Corporation. Many firms delay their pricing until United States Steel

publishes its price list and then adopt its figures for their products. Although such pricing is on a strictly voluntary basis, the procedure has resulted in prices as uniform as if there had been an actual agreement among the competing firms to abolish price differentials. The practice of quoting delivered prices from established basing points also leads to uniform prices within an area. This system has been used in the steel industry, with United States Steel again as the leader, and in many other industries. When members of an industry quote the same prices from the same basing points, competition will become a matter of salesmanship rather than of price.

REGULATION OF COMPETITION

Toward the end of the last century the American public began to be alarmed at the growth of large industrial empires and demanded that the government act to preserve free enterprise. Public opinion became so aroused that both major political parties, in their platforms of 1888, asserted their opposition to monopolies and monopolistic practices. It was practically assured that legislation to curb monopolies and unfair trade practices would be passed by Congress regardless of which party won the election.

The Sherman Antitrust Act was passed in 1890 with the object of eliminating monopoly and restraints of competition. Twenty years later it was evident that the law was inadequate for the complex industrial problems which were developing. Cases tried under the act clearly indicated that preventive rather than corrective treatment was required. Consequently, the Clayton Act was passed in 1914 to render illegal certain devices by which concentration of economic power is made possible. In the same year the Federal Trade Commission Act was passed, establishing the Federal Trade Commission as an agency to investigate antitrust cases and complaints of unfair trade practices, and to assist the Department of Justice in prosecution of antitrust suits. The Commission was also empowered to issue orders prohibiting practices considered to be unfair and detrimental to the economy. Since 1914, there have been two important additions to the statutes governing competition, the Robinson-Patman Act of 1936, and the Miller-Tydings Act of 1937. The former represents an attempt to prevent favoritism and discrimination among customers; the latter sanctions state price-maintenance laws, which permit manufacturers to enter into contracts requiring dealers to comply with minimum resale prices established by the manufacturers.

Sherman Act

This law represented an effort to incorporate into the statutes common law precepts regarding restraint of trade. The act declares to be illegal "every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations." It further declares that "every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with foreign nations, shall be deemed guilty of a misdemeanor." Federal courts were opened to the government or individuals seeking redress against those engaged in unfair or monopolistic practices; and penalties, including imprisonment, were provided for violations.

In enforcing the law the Attorney General may bring criminal or civil suit against violators or file condemnation proceedings to seize property owned by a combination engaged in interstate commerce. Criminal actions are employed as a punitive measure and are intended as a deterrent to wrongdoings on the part of individuals. Civil suits are intended to prevent and restrain violations and to restore competitive conditions. Under such civil suits there are three remedial courses available—injunction, dissolution, or consent decrees. By the consent decree, which has become increasingly important in recent years, the defendant accepts the court order in lieu of long and costly trial proceedings, the substance of the decree usually being prepared on the basis of negotiations between the Department of Justice and the defendant. Private suits may also be brought against violators of the law for the purpose of obtaining damages or an injunction against continuation of unfair practices.

Responsibility for initiating actions and for prosecuting cases under the law is vested in the Antitrust Division of the Department of Justice. In actual practice, the division has never had an adequate staff or ample appropriations for its enforcement program. It has furthermore been hampered by inability to obtain accurate factual information on suspected activities as well as competent witnesses willing to testify in court. As a result, the government has been characterized as possessing a "big act but a little stick" with respect to unfair trade practices.

Under the common law all arrangements, of whatever sort, which tend unreasonably to restrain full and free competition are unenforceable. The federal antitrust act outlawed all such arrangements, but since the act does not contain a definition of unfair practices the question soon arose as

to what constitutes a restraint of competition. Consequently, in applying the act, the "rule of reason" became an important consideration and is in some cases the compelling factor governing court decisions. Under the "rule of reason," the antitrust act is interpreted as outlawing only those restraints of trade which are deemed "unreasonable." Such an interpretation has focused attention on the methods for restricting freedom of traders, curtailing output, and unduly enhancing prices.

Before 1911, the courts interpreted the statute quite literally, laying stress on the language which outlawed "every" restraint of trade, and for twenty years after passage of the act the rule of reason was not applied to antitrust cases. Mr. Justice Taft, in the decision on Addystone Pipe case in 1898, held that the rule of reason could apply only in the case of ancillary restraints to competition. Chief Justice White, however, in rendering the decision in the Standard Oil case in 1911, invoked the rule of reason and it has been generally accepted since that time, although with a variety of meanings. The application of the rule has injected a great deal of flexibility into decisions on antitrust cases by impelling a consideration of the motives and purposes of the defendant in each particular case. It has also encouraged the Court to appraise in each case the possible alternatives to existing business organizations and arrangements, for it is recognized that the effects of abolishing existing practices may be more detrimental to the economy than the disadvantages resulting from their continuation.

In 1895, five years after the passage of the Sherman Act, the first case reached the Supreme Court, and the decision rendered in it clearly indicated that many of the prevailing undesirable practices could not be removed through application of the law. In this case the American Sugar Refining Company of New Jersey had purchased refineries in Philadelphia and had obtained control of about 98 per cent of the sugar-refining capacity of the country. The Sugar Trust had been formed through transfers of stock which brought under one consolidated company a practical monopoly of the industry. Production was controlled and prices had been increased. Nevertheless, the Court ruled, in effect, that a monopoly in manufacture of sugar does not constitute a monopoly in interstate commerce and therefore is not within the scope of the Sherman Act. Interstate commerce was narrowly conceived to exclude activities not directly associated with the physical movement of goods over state lines. This decision meant that manufacturing and mining operations were practically exempt from the act and indicated that, so long as this interpretation prevailed, industrial mergers were not subject to prosecution. Not until Theodore Roosevelt

initiated a "trust-busting" campaign did it appear that the act might become a weapon against large industrial combinations.

Among the early cases attracting considerable attention, the Northern Securities case (1904), the Standard Oil case (1911), and the American Tobacco Company case (1911) were probably most significant. The Northern Securities Company case concerned a holding company formed to combine two parallel and supposedly competing railroads—the Great Northern and the Northern Pacific. The decision in this case ordered the dissolution of the holding company and represented the first time that a holding company was considered to be an illegal device for restraining competition.

The Standard Oil Company had obtained a virtual monopoly of distribution and marketing of petroleum products at the time its case was presented. Although the company had no monopoly over production of crude oil, it had been able to obtain control over from 85 to 90 per cent of the refining business by directing distribution through monopoly of pipe lines and discriminatory freight rates. Impartial study of this case clearly indicates that the dominance of Standard Oil did not result from efficient management and economic operation, but rather from its predatory practices, freight discrimination, and other unfair methods of competition. The record plainly showed intent to capture control of the industry. The Court declared that the methods of obtaining control were illegal, condemned the monopoly, and ordered dissolution of the company. The decision was concerned more with the intent to monopolize than with the actual monopoly conditions per se.

The American Tobacco Company case was in many respects similar, although the actual methods used to gain monopoly power were different. The Court again issued a decree ordering dissolution. Whereas in the earlier days emphasis had been laid on a literal interpretation of the law and on whether or not an arrangement affecting interstate commerce had been formed, the tendency now was to stress unfair practices and application of the rule of reason. Thus the way was being opened for passage of the Clayton Act.

The United States Steel case (1920) resulted in an important decision although it was decided by the margin of only one vote. The corporation —a holding company—had been organized in 1901 to acquire stock in 12 operating companies. The combination was vertical, covering all stages of steel production. The Court found that one of the purposes of the combination was monopolization of the industry, but this goal had never been reached and competition had continued to exist at all times. Price competition had been avoided by pooling agreements in the earlier years and later

by mutual understanding with other firms. The company was apparently following a policy of "live and let live"—to the extent that independents were kept in line by agreement. The Court was not influenced to outlaw the combination because the holding-company device had been used and condoned the intent to monopolize because it had failed to materialize. In fact, at the time of the suit the proportion of total output controlled by United States Steel was lower than in earlier years. The Court's decision strongly implied that "mere size" and the existence of powers not used were insufficient grounds for dissolution. Substantial evidence was presented by the company, competitors, and customers to show that vigorous competition actually did exist. It was apparently the view of the Court that the company had seen the error of attempting a monopoly and had mended its ways, and that, although it exercised price leadership to a high degree, no competitor was forced to follow prices quoted by United States Steel. With respect to the size of an individual unit, the Court was convinced that the Sherman Act did not make great size illegal.

The Clayton Act

As the devices by which unfair competition were attained became prominently recognized in the suits under the Sherman Antitrust Act, it became evident that the law should be made more specific. Accordingly, the Clayton Act was passed in 1914 to prevent the use of holding companies and of interlocking directorates as methods of monopolization. The act also specified certain trade practices as unfair and illegal, in particular, price discrimination and tying contracts which bind the buyer to observe practices that tend to diminish competition.

Federal Trade Commission

The Federal Trade Commission was established to investigate complaints of violations of the antitrust act and to assist the Department of Justice in preparing and prosecuting suits and in formulating readjustment plans. The Commission is also responsible for studies of trade practices and for advising Congress on legislation needed to prevent unfair trade practices. Power to issue orders requiring cessation of unfair trade practices is vested in the Commission.

Before 1938, the Commission's activities were mainly concerned with investigation of complaints made by a competitor against discriminatory acts of others. Orders issued to restrain such acts generally had the effect

of protecting the good will of the complainant by stopping an imitation of his brand name or trade-mark. But some orders tended to aid rather than eliminate monopolistic competition, as in the protection of brand names. With the passage of the Wheeler-Lea amendment to the Commission's act in 1938, authority was granted to the Commission to exercise initiative in restraining practices detrimental to the interests of the public. Although in some cases the public only indirectly benefited from earlier decisions, the Commission can now intervene directly in the public interest against such practices as false advertising or adulteration of products.

Robinson-Patman Act

This act, passed in 1936 as an amendment to the Clayton Act, is aimed at the elimination of price discrimination in the sale of commodities. A major factor impelling passage of the act was the development of chain stores and other large distributing organizations that were able to obtain favorable prices because of quantity buying or their strategic position (e.g., geographic location or bottleneck position in processing). The law prohibits differential pricing which impairs competition, except when such differences can be justified on the basis of cost differentials or on the basis of costs resulting from distributing functions performed by buyers. It is unlawful for a buyer engaged in interstate commerce to induce or accept a price differential prohibited by the act. Quantity differentials in price are not outlawed, but the right is reserved to the Federal Trade Commission to specify maximum quantities beyond which discounts are not allowed. The act was passed as an attempt to protect small and independent buyers from abuses such as abnormally large price concessions for quantity purchases, which favor large concerns. To the extent that allowed differentials in price accurately reflect differences in costs of distribution at the higher levels, compliance with the act tends to equalize competitive advantages with respect to price and to focus attention on nonprice competition.

Miller-Tydings Act

The Miller-Tydings Act of 1937 is an amendment to the Sherman Act relating to resale price maintenance. Resale price maintenance exists when dealers are compelled to observe resale prices established by the manufacturer. This Act legalizes resale price-maintenance contracts covering branded merchandise that is sold in interstate commerce and resold within

any state which has a price-maintenance law. A certain brand of tooth paste, for example, cannot be sold by a druggist for less than the retail price prescribed by the manufacturer. The growth of "super markets" and "cut-rate stores" was an important factor leading to passage of the Miller-Tydings Act, for many such retailers failed to adhere to list prices. Particularly for "long-margin" items, prices at retail were sometimes reduced to the point where small independent stores could not afford to engage in price competition. Frequently, well-known branded merchandise was used as "loss leaders" to attract customers to a store.

California, as the leader among the states in adoption of regulatory legislation, first enacted a price-maintenance law in 1931. Other states soon followed until practically all states now have such laws. Despite the widespread adoption of the principle of price maintenance, the federal law was passed with only a bare majority support, being pushed through Congress as a rider to a District of Columbia appropriations bill.

The effects of resale price maintenance on competition and its benefits to the economy are still not accurately determinable. The Federal Trade Commission has suggested that the Miller-Tydings Act is detrimental to the best interests of the public and for this reason has advocated its repeal. But the act does not affect a very high proportion of merchandise—some estimates place the amount at 10 per cent or less. One important result of the law has been encouragement to the introduction of private brands by large retail organizations. These are sold at relatively low prices in competition with branded merchandise held at list price by the act. Another means of avoiding the restrictions of the act has been the energetic sales propaganda designed to create a market for brands for which price maintenance is not mandatory.

SUMMARY

The decisions of the Supreme Court clearly demonstrate that there has never existed a clear determination of what constitutes a violation of the antitrust act. Because of conflicting decisions and unpredictable interpretations by the Court, uncertainty always exists as to whether an actual violation has occurred until the verdict is announced in the specific case. Limits to permissible action seem quite elastic. Although the Sherman Act was unquestionably intended as a weapon against "big business," the attitude of the Court has indicated that large size is not a compelling reason for finding a company guilty of violation. Under the broad and varied interpretations of the act, monopoly can be defined as whatever the Court

decrees to be unlawful. Since 1911 emphasis has been laid more on the *intent* to monopolize or restrain competition and less on the form of organization and extent of control.

Arrangements to lessen competition have generally been considered in terms of price, with little attention devoted to other forms of competition. Price uniformity, however, does not constitute sufficient grounds for a finding of guilt. The Court has progressively turned to consideration of whether a concern has unilateral power to control price. In recent years the Court has abandoned the early idea that large-scale business is synonymous with restraint of competition and apparently gives recognition to the fact that integration may, in some instances, eliminate competitive waste. Results of possible alternatives open to the Court are appraised and arrangements to eliminate unfair practices are now more frequent than dissolution. Dissolution has generally failed to restore competition, especially with respect to price, as it existed prior to combination.

Recent experience in antitrust cases indicates that to establish a violation of the statutes it is necessary to prove that a company possesses the power, through restriction of output or otherwise, to hold prices above the level which would prevail if competition were freely operative, or that a concern has indulged in practices indicative of intent to monopolize.

PRESENT STATUS AND FUTURE POSSIBILITIES

Business organizations in the present economy embrace types ranging from virtually complete monopoly to pure competition. The closest approach to complete monopoly prevails in the field of public utilities, where public franchises are granted to single firms and so preclude the entry of possible competitors. At the other end of the competitive range there still exists pure competition among producers of many agricultural commodities. But by far the major proportion of all business activity is conducted in the intermediate range of oligopoly and monopolistic competition.

From the standpoint of the consuming public the most desirable kind of business relationship is that which encourages the most economical use of resources and which therefore results in a maximum output of goods and services at the lowest possible prices. To achieve this end, monopoly may be necessary in a few industries, duopoly or oligopoly in others, and pure competition in the rest. But in view of the myriad types and degrees of competition in the present-day economy there exists the problem of determining the nature and extent of regulation needed to secure from industry maximum benefits to society as a whole. One possible solution of the

problem would be an attempt to force a larger role for pure competition. This would require decentralization of economic power and the dissolution of large firms. However, experience with enforcement of antitrust acts indicates that dissolution does not generally result in any appreciable increase in competition within an industry. There is still a real question, however, whether the antitrust laws have ever been seriously enforced. The adoption of this policy would also necessitate the abolition of all except public-utility monopolies.

An attempt to restore pure competition would almost certainly generate confusion and the state of competition resulting from it would be only temporary, unless more vigorous and successful efforts were made to preclude a return to monopolistic conditions than have been demonstrated heretofore. From a production standpoint, the economy would be discarding the results of many years of industrial progress and the devices by which American industry achieved its present high state of development; uneconomic use of resources in some industries and some loss of the economies of large-scale business would ensue.

A possible alternative to attempted restoration of universal free competition is extensive governmental regulation of all phases of competition. This policy is to be criticized for several reasons. (1) Regimentation is unnecessary in large sectors of the economy—even in many of those dominated by large firms—because competition is still a most efficacious influence. (2) Comprehensive control by government is usually inflexible; even if some competitive gains were made by such control they would be partly offset by its inflexibility. (3) Principally, comprehensive regulation involves centralized control and allocation of economic resources, regulation of production and distribution, and dictation of prices. This has been attempted in the United States only during war when conditions are most propitious for success. But wartime experience in this country—and comprehensive planning ventures in other countries—give little reason to hope that a more economic allocation of resources is to be expected from this course. Judged by most standards—aggregate production, per capita production, average weekly earnings or hourly wages—the American economy has done well in relation to its resources and historical development. The market process as the guide to resource allocation is both flexible and relatively efficient.

In a period of unemployment and low levels of national output, the restrictive aspects of monopoly are intensified and public attention tends to be directed toward monopoly problems. On the other hand, growth and expansion in the economy either minimize restrictive practices or encourage new business techniques and competitive behavior. The fostering

of expansion, by creating conditions conducive to new enterprise and investment, may indeed mitigate the problems of monopoly, although it will not warrant a relaxation of antimonopoly policy. Monopoly is a hardy growth; virtually all sellers would like its protective advantages for themselves. Omnipotence on the part of government would be required to eradicate it or to regulate its activity completely. Long-run expansion and dynamic changes in the economy, coupled with constant vigilance against its excesses, may be expected to mitigate the influence of monopoly in a manner most consistent with the character of the economy itself.

CHAPTER 28

The Supply of Labor and Labor Organizations

Gainful workers—those who work, or seek work, for a money income—in 1946 numbered approximately 60 million out of a total population of 141 million. Some gainful workers work for themselves, and the rest work for others and receive wages and salaries. Among those who are self-employed as independent entrepreneurs are farmers and small businessmen (retailers, service station operators, building contractors of many types, small shop owners, etc.), professional persons (doctors, lawyers, and accountants), and a miscellaneous group of odd-job workers. These number possibly as many as 15 to 20 million. The remaining 40 to 45 million in 1946 worked for or sought work from employers—private corporations, government, and independent entrepreneurs. These two groups are not wholly distinct, since individuals constantly shift from one to the other.

Labor organizations confine their efforts to those who work for employers, principally wage earners. In 1946, about 15 million persons were members of unions, about 25 per cent of the total number of gainful workers (nearly 40 per cent of the wage and salary earners). The largest groups of workers represented in union organizations include: (1) the building trades and allied workers; (2) railroad workers; (3) truck drivers

(teamsters); (4) other skilled groups such as bakers, printers, air-line pilots, movie actors, barbers; (5) large industry groups such as coal miners, and wage earners in the automobile, rubber, steel, farm machinery, electrical machinery, and other mass production industries. In recent times membership has grown among low-salaried white-collar workers.

One hundred years ago the union movement was of no consequence. At the present time unions flourish with large memberships, with combined income from dues exceeding 100 million dollars annually, and with total assets many times this amount. The primary purpose of a union is to secure control over a given part of the labor supply—either certain skilled groups, or groups of workers employed in certain industries—in order to increase wages and to obtain other working advantages for its membership. Unions influence the supply, not of labor as a whole, but of the types of labor they control. Thus the supply of labor and the activities of unions are closely interrelated.

COMPOSITION AND CHANGE OF POPULATION

The total supply of personal services is affected by many influences; the size and composition (sex, age groups, etc.) of the population; the willingness and ability of people of working age to take employment; the length of time per day, per week, per year that a person is willing to work. While the effect of some of these influences has been considered in general terms in Chapter 3, they can now be discussed in detail.

The size as well as the composition of population is determined by the ratio and variation of several elements. For one thing, the ratio of females to males in the American population has steadily increased. In 1910, the ratio of males to females was 106.0 to 100.0. The predominance of men was caused in part by the fact that during the great immigration of the decade preceding, single men exceeded single women and many married men were in the country for a time without their families. By 1940, the ratio had declined to 100.7 males to 100.0 females, partly because the preceding period was one of low immigration. By 1950, it is expected that females will outnumber males. The sex ratio does not, from the point of view of the labor force, exercise so great an effect as formerly, because of the entrance of increasing numbers of women into industry.

Age composition, which is the proportion of people in different age groups within a population, greatly affects the size of the labor force. From 1890 to 1940, the median age of the population of the United States rose from 21.4 years to 29.0 years. Along with this there has been a marked decline in the

proportion of the population under the age of 20, and an increase in the proportions of each of the higher aged groups. In 1850, the group under 20 years of age constituted more than half of the population; in 1940, it was 34.5 per cent of the total population; and one prediction expects only 25 per cent under 20 years of age in 1980.[1] In 1850, less than 5 per cent of the American population was over 60 years of age; but by 1940, this group constituted 10.5 per cent. This group is expected to increase as time goes on, although the increase in birth rates in this country during the 1940's will modify prewar population forecasts.[2]

There has been considerable speculation as to the effects of an older average age on enterprise, risk-taking, consumer demand, and the labor force. An older labor force may mean greater stability in labor supply; and this, though desirable from some points of view, may also reduce the labor mobility required by an expanding economy. Moreover, the inflexibility that accompanies age and the unwillingness of old people to venture forth may set up stagnant labor pools at points where the demand for labor is declining, with the result that expanding firms in other areas are deprived of potential labor.

The population of the United States in 1880 was 70 per cent rural and 30 per cent urban. At present over half of the population resides in cities and towns (over 2,500 population). Urbanization accompanied the growth of industrialization; the migration of rural population to the cities built up labor supply available for industry. This shift of labor supply tended to raise incomes in agricultural regions and moderated the increase of wages in urban areas.

In common with the rest of the world, the population of the United States has steadily risen. In 1790 population was 3,929,000; in 1900, 76,129,000; in 1946, 141,000,000. The excess of births over deaths and net immigration are the causes of growth. The birth rate is often measured merely by the number of births per thousand members of the population, and this calculation is called a crude birth rate because it is not refined for age, sex, occupation, or income, or for any other condition which many influence birth and death rates. Given the total number of births and deaths and an excess of the former over the latter, the difference is the natural increase. For example, if the crude birth rate is 14 per cent and the crude death rate is 10 per cent, 4 per cent is the natural increase. If the percentages were reversed, 4 per cent would constitute the natural decrease.

A progressive decline in the birth rate is a trend common to virtually

[1] Paul H. Landis, *Population Problems*, American Book Co., 1943, pp. 278-79.
[2] See Chap. 3, pp. 77-79.

every European country and the United States. Births per 1,000 population have been steadily declining in the United States since 1870. In the years 1871-75, births per 1,000 of population averaged 37.0; by 1920 the rate had fallen to 23.7; it reached a low point of 16.6 at the bottom of the depression in 1933 and rose to 20.9 in 1942. While the birth rate is affected by depressions, prosperity, and wars, a long-run downward movement of the birth rate has prevailed in this country and western Europe. This downward trend has led to a shift in emphasis from limitation of population—a doctrine highly regarded by nineteenth-century economists—to means of expanding it.

At best the crude birth rate is an inadequate measure of human fertility or the actual reproductive rate. Consequently other methods are needed to estimate the effect of changes in age composition on future changes in population. Several measures have been devised, one of which is the "net reproduction rate." The net reproduction rate measures the number of daughters that would be born to 100 women during their lifetime assuming the continuation of existing birth and death rates by specific ages. For example, a rate of 100 indicates that 100 daughters are born to the first group. Thus there would be enough to replace the original mothers, and the population would remain stationary. A higher rate would indicate some increase in population, a lower rate a decline, provided that existing birth and death rates continue. The net reproduction rate in the United States has been declining. Figures between 1920 and 1940 show that it dropped from 116 in 1920 to 94 in 1933 but rose to 96 in 1940.

The growth of population in the United States owes much to immigration. Net immigration reached an annual average of 251,000 in the decade preceding the War between the States. In the 1880's, annual net immigration more than doubled and in the first decade of this century, it reached an annual rate of 880,000. Before the 1880's the great majority of new arrivals came from the northern European countries and the British Isles; in the 1880's and until the restriction of immigration during the 1920's the great majority came from southern European and the western and eastern Slavic nations. Without the millions of willing hands American industrial expansion would undoubtedly have been retarded. The readiness of immigrants to work at low wages led businessmen to promote immigration. Various objections had been raised many times, however, against free entry of foreigners into the United States. Objections were based on cultural, educational, and nationalistic grounds. Organized labor always opposed immigration on economic grounds, for the constant increase in the supply of labor tended to lower wages or to hold down the rate of increase. As immigrants flocked to the unskilled occupations, the disparity of wages

and standards between the skilled and the unskilled was sharply accentuated. The availability of a large amount of cheap labor enabled organized skilled workers to widen the difference between wages of skilled and unskilled to a greater extent than was possible in England or on the European continent. Not only did immigrants tend to depress wages, but the constant increase in the supply of labor may have slowed the rate of technical progress, for the advantage of substituting capital for labor is not high when large masses of unskilled, low-paid workers are available.

After World War I the restriction of immigrants from Europe was undertaken largely because of (1) the opposition of organized labor (on economic grounds) and of certain industrial and agricultural businesses, (2) the opposition of groups for cultural and national reasons who feared the "dilution" of the Anglo-Saxon stock, and (3) the fear of business groups —formerly proponents of immigration—that immigration encouraged the importation of radical doctrines. Thus Congress enacted in the 1920's the immigration quota laws, culminating a movement to limit immigration that had gained strength during the preceding century. Annual immigration from Europe was limited to 153,774, and the total was divided among the applicants from other countries in accordance with the percentage of peoples of the same origins already represented in the population of the United States as revealed by the census of 1920.

The restriction of immigration forced employers to seek new sources of common labor. Mexicans and rural Negro and white workers were recruited to fill the gaps left by the reduction in immigration. Labor agents invaded Southern farming districts; and in self-defense these communities sought to limit and even prevent the depletion of the local labor force. Little opposition was encountered in recruiting Mexicans (not included by quota restriction), and railroads, steel and metal industries, and growers of some agricultural crops found the Mexican workers a helpful substitute for the former immigrants. While American industry profited from the presence of these workers, this gain was purchased at a social cost to both the workers and the whole society. Competition on the labor market aroused the hatred of white labor, and deterioration of real-estate values aroused the opposition of real-estate owners to the new workers. The race riots which occurred after World War I were in part due to the animosity against the increase of colored workers in many communities.

Restriction of immigration has also had other effects. The shortage of supply of certain types of labor (e.g., tailors at the present time) may be traced to this policy. Wage increases have been more readily attained by wage-earning groups freed from the competition of immigrant workers.

In the long term, mechanization of industry may be stimulated because of the absence of cheap immigrant labor. The sharp reduction in immigration will serve to eliminate or reduce racial and national differences within the labor force. A labor force homogeneous in language and social mores is no doubt one which is easier to organize in labor unions.

HOURS OF WORK

The supply of labor service is determined not only by the size of the labor force but also by the number of hours workers are ready to work. The hours worked per day and per week in American industry have been steadily reduced since the early nineteenth century. From almost the beginning of this nation's history, demands for shorter hours have been made by workers. As might be expected, the working-hour standards, or customs, prevailing in agriculture also prevailed in industry at the outset; subsequently, the drastic differences in the conditions of employment between agricultural and industrial areas stimulated organized demands for shorter hours by industrial workers. The working regimen on a farm is looser and the discipline is more relaxed than in a factory or construction job. Punctuality and uninterrupted work are requisites in industrial employment. Because workers considered industrial labor more onerous, they made demands for 12-hour days, then 10, 9, and shorter days. As the workday was reduced, the demands came to be based on productivity and the need and desire for more leisure. Studies conducted during both world wars indicate that for given tasks there is an optimum workday and work week. Beyond a point output per worker does not increase, and eventually decreases as hours are lengthened.

Legislation over the last century has been enacted at times to restrict the number of hours that could be worked per day or week. At the outset these restrictions affected minors and women, and were based upon considerations of health and welfare. Reduction of the hours of labor of men has been brought about by trade-union action, by voluntary concessions of industry, and by legislative regulation in certain trades held to be dangerous to workers or to the public safety.

Hours worked per day only a century ago appear unconscionably long by present-day standards. A hundred years ago, the typical workday in American industry was 11½ hours, and the work week was 6 days. Even in 1890, blast-furnace men in steel mills worked 7 days of 12 hours each, or 84 hours per week. From the very outset of the shorter hour movement the building trades were leaders, and by 1840 they had established the 10-hour workday in the large eastern cities. Factory workers were not so well situated, for

the majority were working 11 and 12 hours per day. By 1890, the building trades were working 52- or 54-hour weeks, and factory workers were very largely working 60 hours per week.

The movement for the 8-hour day spearheaded by the American Federation of Labor in the 1880's was partly responsible for reductions in the hours of labor. Between 1890 and 1929, the hours of labor decreased. Weekly hours of building tradesmen fell from an average of 51.3 in 1890 to 43.1 in 1910; of coal miners from 60.0 to 48.5 hours in the same period. In manufacturing, hours were reduced from about 60 in 1890 to 50 in 1929. The depression of the 1930's led to sharing of the work on fairly large scale; although the scheduled hours of work per week were not drastically changed, workers were frequently employed only a small fraction of that time. The enactment of the National Industrial Recovery Act in 1933 gave official support to the movement to spread the work by shortening hours. In the codes of fair competition provided for in this act the hours of labor were curtailed; the 44-hour week was established in most industries, and the 35- or 36-hour week in a number of others.

Under the Fair Labor Standards Act of 1938, work after 40 hours per week—in interstate industries—must be compensated at one and one half times the regular rate. The penalty rate on overtime is designed to prevent the expansion of the work week. This act was largely depression-inspired, namely, to spread employment. The Fair Labor Standards Act was the culmination of a long-term movement for shorter hours which reduced the work week by nearly 50 per cent in industry over the last hundred years. Both the supply of labor and the productivity of labor have greatly increased during the same period, thus permitting increased output while at the same time hours have been reduced.

THE LABOR MOVEMENT

The term *labor movement* has been applied to the activities of unions of wage earners to improve conditions of employment The labor movement began in the United States early in the nineteenth century although the first union was organized among the shoemakers of Philadelphia in 1792. In the 1820's local organizations of skilled workers were founded to press for higher wages, shorter hours, and educational and political reforms. These efforts mark the first attempts of labor in this country to organize as a class or group. They were sporadic in character and were usually met with determined opposition from employers, the public generally, and the courts. During the 1850's the modern union movement had its inception

with the advent of *national* unions. A national union is comprised of local unions in different parts of the country. The locals are subordinate to the national union in somewhat the same sense that a chapter of a fraternity is to the national organization. In the course of time many of the national unions established locals in Canada, and therefore were known as "international" unions. These early organizations became casualties to political ambitions, employer opposition, the courts, or to economic depression.

In the 1860's union organizations grew rapidly as international unions were formed in many trades. The causes of this revival are to be found in the economic development of the country in this period. The growth of industry and transportation created a national market for a large part of total commodity output. With labor costs an important part of total costs, the lower wage of one employer gave him an advantage over the high-wage employer in competing on the national markets. Consequently, low wages at one point stimulated efforts to bring about low wages at other points. Even when low wages at one point did not cause employers at another point to seek a lowering of wages, they retarded increases in rates because high-wage employers could adduce the possible deterioration of their competitive position as an argument against additional wage increases. These economic influences stimulated the growth of unions as means of maintaining and increasing wages over a wide market. Printers, molders, machinists, tailors, bricklayers, and locomotive engineers were a few of the trades in which national organizations arose in the 1860's. The process of establishing national unions did not take place at the same time and at the same rate in all trades, for progress was influenced by a variety of factors not common to all branches of industry.

National Central Organizations

In the period following the War between the States the first two attempts to combine several national unions into a combined organization failed. The National Labor Union, organized in 1866, represented the first effort to achieve national unity of labor. This movement attracted a wide variety of groups, a number of which, including socialists and advocates of inflationary monetary policies, were largely interested in political action. While labor frequently engages in political pressure, the history of the American labor movement is studded with examples of failure attributable to the shift of interest from immediate economic gains for its members to the promotion of political programs. Greenbackism—the inflationary monetary policy—

broke up the National Labor Union. Virtually all workers in a shop or a trade will approve of and work for higher wages and better working conditions, but differences are likely to arise as soon as political issues are given a dominant place in the program. As soon as the National Labor Union became absorbed in politics, the trade unionists withdrew. The political reformers lacked the persistence and support to keep the organization going. It expired in 1872 after naming a candidate for the Presidency of the United States who declined the nomination.

For a time it appeared that the "Noble Order of the Knights of Labor" might become the means of uniting the diverse labor organizations into a united and powerful whole. Started in 1869 as a secret organization by a small group in Philadelphia, the organization at first progressed slowly. In 1870 the locals, called local assemblies, admitted "sojourners," i.e., workers from other trades. The units of organization were the local assembly, made up of wage earners to the extent of at least three quarters of the membership; the mixed assembly, made up of workers from several crafts; and trade assemblies, containing workers of only one trade. District assemblies were made up of delegates from both trade and mixed assemblies in a city or state. There were also national trade districts which were in fact national craft unions (representing workers in one craft) or industrial unions (representing all the workers in an industry). The governing body was the general assembly, or national convention, and at the head of the organization stood the Grand Master Workman. In 1878 the order abolished secrecy, and for the next ten years underwent a spectacular expansion of its membership and prestige under the leadership of Terence V. Powderly. Strikes against the commercial telegraph companies, several railroads, and the Chicago meat packers were led by the Knights in the 1880's. But defeats, especially in the latter two struggles, lowered their prestige and weakened their membership.

While the Knights of Labor were expanding, trade unions continued to grow on a national basis. Although some of the trade unions affiliated with the Knights, many remained outside its ranks. Many trade-union leaders opposed the intrusion of the Knights into trade-union activities since membership in the Knights was not limited to wage earners or were its aims the same as trade-union aims. Although the Knights conducted the longest strikes in the 1880's, a number of its leaders were basically hostile to trade-union action. They had less faith in strikes, boycotts, and other bargaining weapons than in producers' co-operation and political action. Here again political programs brought about a cleavage in the labor movement.

The American Federation of Labor

Even during the heyday of the Knights of Labor, the trade unions sought to combine for common aid and defense. In 1881, the Federation of Organized Trades and Labor Unions of the United States and Canada was formed. It took an important step toward the present form of labor organization by making national unions its basis and by stressing its primarily economic purposes. Nevertheless, national unions did not make very rapid progress, partly because jurisdictional disputes between the Knights and the trade unions impeded their progress. In the spring of 1886, the heads of a number of trade unions proposed a treaty to the Knights which would have in fact barred the latter from participating in trade-union activity. The Knights rejected the conditions, and the trade unionists called a convention for December, 1886, at Columbus, Ohio. This convention launched the American Federation of Labor, with which the Federation of Organized Trades and Labor Unions was merged.

The Knights of Labor reached its peak membership in 1886, but following this time it steadily lost membership and influence. In contrast, the American Federation of Labor, which gained members only slowly at first, by the turn of the century became the representative of the largest proportion of the organized workers of America.

World War I furnished a very favorable opportunity for labor organization. The wartime shortages of workers and need for increased production, as well as a favorably disposed national administration, all aided in the unprecedented growth of union organization. Its expansion temporarily was stopped by an employer antiunion campaign in the early 1920's launched under the name of the "American Plan," and by the conditions of postwar depression in 1921. Throughout the 1920's, labor organization could not regain its lost momentum; despite the prosperity of the period, labor unions not only failed to increase membership but in fact actually lost members.

The American Federation of Labor is a federation of independent and autonomous national and international unions. In 1946 it included more than 100 unions and about 7,500,000 members. Each union in the federation is independent; it can set its own rules with respect to admission, dues, and initiation fees. The American Federation of Labor is dominated by the international unions which control the majority of votes at conventions. These unions are made up usually of members of a single craft, such as carpenters or teamsters, or of closely related crafts.

Major Changes During the 1930's

The fortunes of organized labor took a turn for the better under the New Deal administration. The large inroads into union membership made in the 1920's and during the depression years of 1930-32 were ended, and union activity began to flourish again. Under section 7a of the National Industrial Recovery Act, labor was guaranteed the right to organize and to bargain collectively in unions of its own choosing, free from interference or coercion by employers. Established union membership as well as new unions grew rapidly, not only in old trades and industries but also in the newer mass production industries such as automobile and rubber manufacturing.

As soon as unions in the mass production industries began to form, the American Federation of Labor was confronted by a jurisdictional problem. A "union jurisdiction" refers to the type or class of workers eligible to join a specific union. The employees in mass production industries were largely semiskilled production workers who had never belonged to a union, and many were either unsympathetic or apathetic toward unionization. Since the older unions in the American Federation of Labor were largely made up of craft and semi-industrial unions, they demanded that the workers in the mass production industries be divided by crafts and assigned to existing unions. Thus the electricians' union in the American Federation of Labor demanded the membership of electricians employed in steel mills, in the automobile plants, in the aluminum plants. Similar demands were made by the machinists' union, the patternmakers' union, the carpenters' union, and other affiliates of AF of L. On the other hand, many union leaders in the mass production industries believed more effective organization would be achieved if all who were engaged in work at a large plant would, irrespective of standards or grades of skill, join one union composed of all workers in the industry of which the plant was a part.

While the question of the form of union to be organized was the main difference at this time, other issues played a role. The American Federation of Labor had never concerned itself with large-scale organizing of workers; this function was generally performed by its affiliated international unions. The leaders of these large unions were at the time skeptical of establishing permanent unions in the mass production industries. Their skepticism was based on their lack of success in the past in rallying such workers; moreover, they felt that unionization in a particular area should not depend too directly upon outside organizers and should be accomplished by local

units of craft unions. In general, the AF of L was not accustomed to organizing on a large scale and at this time had neither funds nor personnel to accomplish it.

Compromises of the differences that arose between the newly organized mass production unions and the older unions in the American Federation of Labor were proposed; however, since the federation itself was controlled by the older unions and the new unions were aggressive and unyielding, a compromise was difficult. The new unions in the AF of L formed the Committee for Industrial Organization under the leadership of eight powerful unions (including the miners, the garment workers, and the textile workers) for the purpose of spreading the doctrines of industrial unionism. Ordered to disband by the American Federation of Labor, the unions of the CIO were suspended from the AF of L when they refused to do so. Under aggressive leadership and with generous financial support furnished by the mineworkers and other dissident AF of L unions, the CIO unions spread their influence and, after a number of spectacular strikes, deeply entrenched themselves in the American economy. They later set themselves up on a permanent basis as the Congress of Industrial Organizations. In 1946 the membership of the CIO unions was about 6 million.

STRUCTURE AND FUNCTION OF LABOR ORGANIZATIONS

American labor unions, whether independent or affiliated either with the American Federation of Labor or the Congress of Industrial Organizations, are generally similar in structure and function. Union organizations at present may be classified into four groups: (1) affiliates of the American Federation of Labor; (2) affiliates of the Congress of Industrial Organizations; (3) nonaffiliated national and international unions, such as the Brotherhood of Locomotive Engineers and the Brotherhood of Utility Workers; (4) and unions limited to one plant or to one company which are unaffiliated with any other group of organized workers.

Unions unaffiliated with any group outside the plant or company are usually single-plant or single-company industrial unions. Such a union may be a successor to a former company union (a union formed by and usually controlled by the employer, a practice now prohibited by law) or a union not affiliated because the workers may be hostile to outside affiliation.

The make-up of the constituent branches of national unions is generally similar whether or not they are affiliated with another organization. The nonaffiliated railway brotherhoods, for example, are similar in structure to

the carpenters of the AF of L and steel workers in the CIO. The basic unit of organization is the local union. A local of a *craft union* is made up of workers in the same craft or trade, as in a bricklayers' union. A second type is the *amalgamated union,* which is perhaps better described as a quasi-industrial union. This type organizes most—but not always all—of the workers in a plant. The machinists' union is an example, for in some machine shops it may organize all but a few members of other crafts. The *industrial union* is made up of all production workers in the industry. The steelworkers' union and the automobile-workers' union, both CIO affiliates, are examples of this type.

A local may be organized in a single plant. Unions that have a large number of members in a locality may find it desirable to form such locals, called shop locals. Or a single local may be formed in a city and its environs. Some unions allow more than one local to exist in one community. As an example, the teamsters will organize a local of coal drivers, another of food drivers, a third of milk drivers. In smaller communities there may be only one general local, including all crafts. Local unions, in accordance with their charters, collect dues, negotiate contracts, regulate admission to the union, and protect the interests of the union. Locals are administered by salaried officers normally elected by the membership. Paid officers of a local are usually the business agents or business representatives who not only aid in the negotiating of agreements but also help in the adjustment of grievances and in the maintenance of discipline.

Some unions have district organizations, usually called district councils, covering a larger geographical area than the locals; as a rule a number of locals will make up a council. The councils co-ordinate and aid the locals in the performance of their functions and try to promote a common-area policy.

The most important organizations in the American labor movement are the international unions, most of which are affiliated with either the AF of L or the CIO. Local unions are subordinate to the international, and the latter is charged with enforcing the rules and obligations of the union upon subordinate groups. International unions in many instances control the strike fund as well as the funds for fraternal benefits—sickness, old age, unemployment, or death—where these benefits are offered. In addition, most union constitutions make the international union the guardian of the interests and welfare of the organization, and the national officers can, for the good of the organization, revoke the charters of local unions. Sometimes the international sets up rules for admission, minimum and maximum dues,

procedures for electing officers, etc., which subordinate locals must accept. International unions frequently have the power to call and settle strikes. Many unions require that a strike by a local be approved by the international before it can be called. Internationals are also charged with auditing the accounts of locals, checking complaints against local officers, and hearing appeals of members penalized for alleged violations of union rules. Internationals are themselves autonomous; i.e., they are independent and they make their own rules and regulations with respect to their internal affairs.

International unions are governed by conventions which meet at specified periods of time. Legislation enacted at conventions becomes the law of the union, although some unions require that changes in the constitution must be approved by referendum vote. Between conventions the international is administered by its president and executive board.

The salaries of local officers are usually slightly above the amounts earned by the members in the union. Salaries of general officers are much higher. A study of the salaries of presidents of 62 international unions showed that salaries ranged from $2,000 to $30,000 per year.[3]

In addition to the structural forms of unions described above, labor organizations have formed local trades councils. These are made up of representatives from a number of unions whose members all work in the same industry. A good example is the building-trades council to which all the unions engaged in the building trades send delegates. Trades councils also exist in the metal trades and on the railroads; the latter are called system federations and cover an entire railroad or group of roads under the same management.

The American Federation of Labor has also organized several departments. They are made up of the international unions in a single industry and were set up in part for the reasons that impelled organization of the CIO. It is their function to promote common action, co-operation among the unions, and, if possible, to settle jurisdictional disputes.

The AF of L and the CIO are the two central organizations, and both are largely concerned with promoting the interests of labor through legislative lobbying, membership drives, and public relations. Both organizations carry on considerable publicity, and have research organizations to investigate issues that may affect their members. The CIO and recently the AF of L have established special departments for the purpose of organizing workers who presently do not belong to unions.

[3] Philip Taft, "Understanding Union Administration," *Harvard Business Review*, Winter, 1946, p. 256.

Finances of Unions

Unions receive their incomes from the following sources: dues, initiation fees, assessments, charter fees, reinstatement fees, and income from securities they have purchased. By far the most important source is the regular monthly dues. An examination of 250 local union constitutions showed that dues in unions of the AF of L ranged from 75 cents to $7 per month. The most frequent amount was $2, charged by 68 locals; $3 charged by 39 locals was second in frequency.[4] The great majority of these unions belonged to the skilled building and printing trades, which pay death and other fraternal benefits. Dues in the unions of the CIO are appreciably lower. A majority of CIO unions charge $1.50 or less per month. The lower rates are partly explained by the absence of fraternal benefits, and in part to the large-scale and extensive organizing drives made by these organizations. There are apparently economies of large scale even in labor unions.

An examination of three hundred union constitutions showed that initiation fees ranged in AF of L unions from $2 charged by five locals to $350 charged by one local. The most common fee was $50, charged by 76 locals, with $25, charged by 56 locals, the second most frequent, and $100 by 38 locals, third. As is true of dues, initiation fees are lower in CIO unions, most of which charge $10 or less as an admission fee. Despite impressions to the contrary, unions do not readily levy assessments, for these are irregular contributions and are often a cause of dissension.[5] Most of the dues received are as a rule retained by the local union, which pays to its international a specified amount per member per month. This is called a per capita tax, and it is used to finance all of the international union's activities and to build up a reserve.

Union Discipline

Unions insist that their members obey certain rules, and prescribe penalties for violation. Union constitutions may describe the punishable offenses, the manner of trial, and the appeals that can be made to appellate union tribunals. When a union member charged with some offense against the rules has used up all of his rights within a union, he may under certain circumstances appeal to the courts for redress, although the courts frequently will refuse to intervene. Union discipline is often difficult for officers to

[4] Philip Taft, "Dues and Initiation Fees in Labor Unions," *Quarterly Journal of Economics*, February, 1946, pp. 219-32.

[5] Taft, *op. cit.*

maintain, especially in new unions. This problem relates not merely to infractions of union and working rules by members, but also to local officials in their relationship with district and national officials of the union. Unions, like many business, social, and political organizations, have their internal bureaucratic clashes and rivalries. These not only disturb internal relationships but also sometimes good relationships with employers.

In a larger sense, the problem of discipline is involved in jurisdictional disputes. These disputes are more common among the well-established unions of the AF of L. Two AF of L unions in, say, the construction field may both claim jurisdiction over a certain part of the construction or installation work. The employer may be willing to deal with either one or with both, but each claims exclusive jurisdiction. If either one is given the job by the employer the other one may call a strike. Jurisdictional disputes are often settled by agreement in favor of one union; occasionally the employer must in effect hire both unions to do the work.

Jurisdictional disputes may arise as a result of changes in materials (e.g., metal doors instead of wooden doors), methods of work, or the effort of a stronger union to invade the jurisdiction of a weaker one (a practice called union imperialism). Jurisdictional disputes have arisen as a result of the rivalry among the two major national labor organizations. The AF of L and CIO might attempt to organize the same industry or plant, or each might seek to enlist an independent union or capture membership from the other in a given plant. In the ensuing conflict the employer and the public may be damaged or inconvenienced. The rivalry among different unions, and even among locals in the same national union, is sometimes fully as intense as the rivalry among competing business firms.

Union Labor and Nonunion Labor

Union membership is by no means universal in the United States. Membership totaled about 15 million persons in 1946, covering principally the skilled trades and wage earners in the large manufacturing plants. This number is about two fifths the total employees working for wages and salaries who numbered more than 40 million in 1946. Organized labor is nevertheless a potent if not an entirely cohesive minority. To speak of "labor" as a single group is no more realistic than to speak of "business" as a group, or "farmers" as a group. Of course, on certain issues most businessmen may react in more or less the same way; so also do wage earners or union members. Among wage earners, as among businessmen, market interests may clash. Jurisdictional strikes are an example. Similarly, if one

union, such as the coal miners, goes on strike causing great numbers of other union (and nonunion) workers to lose their jobs and income, the latter may regard the strike with ill grace.

The cohesiveness suggested by the general appellation "labor" is weakened, too, by disaffection among wage earners. Some resist membership even though most of their fellow workers in the occupation are members. And some join unions because they regard the union membership card as a "work permit." The rapid loss of membership in a union on certain occasions, a few of which have been mentioned in the history given earlier, reflects this lack of loyalty on the part of some of the membership.

These points do not suggest a weakening of organized labor. On the contrary, union organization seems destined to grow and become stronger, paralleling to an extent the growth and concentration among business organizations. Diversity of interest in an economic sense, personal disinclination to submit to organizational discipline, and organizational (or bureaucratic) conflicts are found in this aspect of economic life as among the entrepreneurial groups.

CHAPTER 29

Labor-Management Relations

Organized labor is interested principally in obtaining increased wages and other advantages for its members by bargaining with employers. Some unions, in addition to promoting better terms from employers, have long-established welfare activities, such as unemployment, old age, sickness, and death benefits as well as social activities. Occasionally unions enter into politics; and some regard the union movement as the vehicle of radicalism. Social and political aims, particularly as expressed by intellectual leaders, are more common in the European labor movement than in the United States, where unions have had a more limited program, one which has been succinctly if not entirely accurately summed up by "more, now." Interested primarily in bargaining, the American union movement has earned the name "business unionism." In their bargains with employers

unions have confined themselves largely to the immediate questions of wages, hours of work, and other conditions of employment. Like all other sellers, or selling agencies, they regard the immediate bargain as the important issue. The larger economic issues affected by their wage bargains are usually given no consideration. Indeed, when they do relate their activities to the general issues of prices, income distribution, and production costs, unions tend to behave usually in the same manner as employers—identifying the general interest with their particular interest.

In bargaining over wages and other aspects of employment, the immediate conflict of interest is apparent; it is substantially the same as the conflict between buyers and sellers in all market transactions. Wages are in general the principal cost of production to the employer paying them. Thus an increase in wage rates is likely to mean immediately a reduction in the employers' income unless the employer can pass the higher wage on to his buyers in the form of a higher price, or offset the wage increase by reducing other costs of production. For employees wages are means of livelihood and well-being, the prices of their services. The wage bargain is therefore simply a wrangle over a price in which both sides exert their respective advantages to the utmost to get the best possible deal. The employer usually has one general advantage, namely, his superior financial position. Unions have sought to offset this by collective action, which enables them to make a fight over wages a more costly matter to the employer.

Union-employer relationships have changed over the decades, and the change has been toward less drastic (for the persons involved) conflicts in disputes. In the present period major strikes come and go without the bloodshed that was not uncommon a few decades ago. The change has come about largely because of legal restrictions placed on employers, forbidding them to carry on antiunion activities and forcing them to bargain with unions, and because employers have progressively become more willing to bargain collectively. Employer tactics have generally changed from recalcitrance and opposition to practicing the arts of persuasion around the bargaining table. The tactics of unions have changed also in the course of their development, but unions still have available a number of weapons for use during a contest of strength.

LABOR WEAPONS

Strikes

A union's principal weapon is the strike, which is an organized withdrawal of workers from their jobs for the purpose of exercising economic

pressure upon an employer to force concessions from him. The simplest kind of strike is against the single employer and may be conducted by a single craft or by all the workers employed in one plant. Interfirm strikes are strikes affecting a number of firms engaged in the same industry. If enough firms are affected such strikes take on the aspect of general strikes (defined below). Sympathy strikes are called not to redress a grievance, but in support of or in sympathy with other workers who claim a grievance. In the building trades, no craft will remain on the job while one of the other crafts is on strike. The refusal of workers to cross picket lines is an example of a sympathy strike. Sit-down strikes are strikes in which the employees cease work but remain at their places of employment.[1] "Quickie" strikes are usually called suddenly without prior notice or warning, being designed to paralyze production quickly and so force negotiations. A "slowdown" strike is a demonstration in which employees remain on the job but perform less than their normal quantity of work.

General strikes may involve an entire community or even an entire nation. They involve a suspension of work by the entire body of organized labor in the area, excluding customarily workers who perform services vital to the public health and safety. General strikes usually have political aims—pressure for some governmental change either in policies affecting organized labor or in the political make-up of the government. A number of national general strikes have occurred in Europe, the most important being the British general strike in 1926. This strike was called in support of the coal miners because the majority of British labor leaders and workers were convinced that the coal owners were seeking to destroy the union and so retard the national labor movement. The strike was lost, and the consequence was the enactment of a law forbidding strikes for political purposes. This law was repealed by the Labor government in March, 1946. No national general strikes have taken place in the United States, or has one ever been seriously contemplated. There have been local general strikes in Seattle in 1919, in San Francisco in 1934, and (briefly) in Oakland in 1946.

The general strike is a dangerous weapon for labor for it is usually so destructive in effect that the government must intervene to stop it. If a general strike lasted any period of time, it would inevitably force authoritarian intervention, for no government can allow a sudden strangulation of community existence without abdicating its power. Consequently labor leaders are reluctant to resort to this extreme weapon.

Strikes cause serious economic losses; this is perhaps the most important

[1] They were widely used during the extensive organizing drives of the CIO in 1937, but as a result of adverse court rulings this technique has been largely discontinued.

argument against a too ready use of this weapon. Strikes can be tolerated in a free economy, but their undesirability need not be gainsaid. Nevertheless, it is sometimes easy to overemphasize the losses resulting from strikes, for the extent of the losses differs from industry to industry and is variously dependent on business conditions. A strike in a seasonal industry might simply mean the shifting of the work season, so that the time lost during the strike will be made up in a normally slack period. In a highly competitive industry a strike against one firm may lead to the placing of orders with competitors, so that the losses incurred by one firm and its workers might be offset by gains to other firms and their workers. However, a strike in an industry producing a perishable commodity and one that cannot be adequately supplied by other firms, e.g., newspapers, will lead to permanent losses. The indirect losses caused by strikes must also be considered. A prolonged strike in the coal industry, or a suspension of streetcar service, leads to severe losses for firms and persons not parties to the dispute.

Picketing

Picketing is a means by which unions enforce and publicize strikes. It consists of workers patrolling before the plant, usually advertising their grievances or opinions by means of banners. It is a means of announcing the existence of a labor dispute and of discouraging or preventing nonstrikers from returning to their jobs. Peaceful picketing is permitted, even though it is designed to exercise pressure upon the employer. Under a decision of the United States Supreme Court in *Thornhill* v. *Alabama* (310 U.S. 88, 1940), peaceful picketing is a lawful form of communication. Workers have a right to announce the existence of a labor dispute and to seek to persuade others not to work for an employer engaged in one. But not all forms of picketing have been recognized by the courts. Mass picketing was enjoined in a number of strikes during 1945 and 1946, and the United States Supreme Court has upheld an injunction prohibiting mass picketing; for while picketing is protected as a form of free communication, the courts may outlaw the type of picketing which promotes violence.

Boycott

The boycott is not a potent weapon but it is still widely used. A boycott is an organized withdrawal of patronage from an employer who is in dispute with his employees. Usually it is accompanied by efforts to induce

other union members and union sympathizers from purchasing the goods of the allegedly unfair employer. It is in fact economic ostracism, to force concessions by nonpatronage. Boycotts may be carried out only by the workers immediately affected—primary boycotts—or they may take the form of persuading third parties not involved in the dispute to withhold patronage—secondary boycotts.[2] Boycotts can be effective in consumer industries patronized heavily by wage earners, e.g., boycott against an overall manufacturer might have fatal results to the employer's business; and firms in other industries might be affected to a lesser extent. Another form of (secondary) boycott is the refusal to handle goods or equipment produced by a nonunion firm. This can be very effective, for firms producing semifinished goods might find that workers employed by their customers refuse to handle their products. While such a boycott is technically feasible and has been used effectively in the past, boycotts have been most successful against producers of consumer goods.

Political Action

Political pressure has always been recognized as a weapon of unions but has also aroused widespread differences among labor leaders. From its beginning, the American Federation of Labor avoided direct political action. The founding fathers of the AF of L and Samuel Gompers, its president for many years, believed—and they had in view predecessor organizations such as the Knights of Labor—that direct political activity by unions would lead to serious division in the ranks of labor. Since the affiliated unions of the AF of L are independent and autonomous, there has always been a minority of unions which urged more active political participation than was practiced by the central organization.

After World War I the hostility of the courts, as well as of state and national governments, forced the AF of L to abjure its neutrality in politics, and in consequence the nonpartisan policy of supporting friends of labor and opposing its enemies was evolved. In the Congressional elections labor campaign committees were usually appointed, and attempts were made to defeat known opponents of unionization. In Presidential campaigns the AF of L has avoided committing itself except in 1924 when it endorsed the candidate of the Progressive party. With the expansion of labor unionism in the 1930's, a number of unions organized Labor's Non-Partisan League, which actively contributed to the Democratic victory in 1936. Much more important was the CIO Political Action Committee, organized in 1943. It

[2] The Taft-Hartley Act of 1947 placed restrictions on secondary boycotts.

was active and powerful in the elections of 1944, but had less success in the Congressional elections of 1946. Many of its supporters, moreover, were nonunionists presumably concerned with promoting labor's economic and political interests.

COLLECTIVE BARGAINING

Collective bargaining is a means whereby the terms of work are determined by the employer or by a group of employers negotiating with a group of employees normally acting through union representatives. Collective bargaining can be better understood by comparing it with individual bargaining. Under the latter condition the individual worker must usually accept the terms of the employer, although the employer is himself not a free agent: he cannot set an arbitrary rate of pay, for his wage offer must take into account conditions in the labor market, area rates of pay, and other relevant matters. Markets are not perfect, however, and the employee negotiating for himself may, because of his own lack of knowledge or the employer's superior bargaining position, strike a poor bargain for himself. The higgling of the market is seldom achieved in individual bargaining, as one may discover by observing the hiring of help at a large unorganized plant. A rate is set, and the applicants must either accept or reject it. Of course rates cannot diverge markedly from the going rates for the type of skills required. Nevertheless, a single employer, especially a large one, has some latitude in the setting of the terms of employment, while the unorganized worker has virtually none. In contrast, collective bargaining implies negotiations, argument, and higgling before agreement over terms and conditions of employment are reached.

As soon as a plant reaches a relatively large size it becomes difficult if not impossible to negotiate an agreement with the entire group of workers. Therefore collective bargaining is usually conducted, at least on the employees' side, through committees representing all workers covered by the negotiations. In a number of industries the employers also act through a representative committee. The simplest form of collective bargaining is the type carried on by the representatives of the wage earners with a single employer. In industries such as building construction, clothing, and local trucking, the employers may combine into an association and deal with the union or unions through an elected committee. Also, regional groups of employers often combine together for the purpose of dealing with the union, as in over-the-road trucking. National collective bargaining exists in the bituminous coal-mining, and railway industries, and in glass-bottle

blowing. Representatives of virtually all the employers meet with a union (or unions) to decide the terms of employment.

The individual worker, although aware of his own needs, lacks a detailed knowledge of business conditions or of area and industrial wage rates. He is not trained or experienced at bargaining. Securing a job is of much more importance to him than is the addition of a single worker to the employer. An employer can protect himself in the employment bargain by providing effective supervision and if need be by fines and deduction from employee wages for spoilage and poor work. On his part, the individual worker often cannot hold the employer strictly to the employment bargain. These disadvantages can to some extent be reduced by collective bargaining, specifically by requiring more precise definition of the reciprocal rights and duties of employers and employees, and by providing procedure for review and appeal of arbitrary action of this nature by the employer or employee.

Coverage of Union Agreements

The culmination of the bargaining process, if agreement has been reached, is a written contract. The spread of union organization in recent years increased the number of workers covered by union agreements to 15 million by 1946. About 70 per cent of all production workers in manufacturing are working under union agreements. Railroad unions cover more than 95 per cent of all workers in the industry. Major companies in steel, automobile production, rubber tire manufacture, aluminum, cement, and shipbuilding industries are unionized. Building construction in the large cities is conducted under union agreements. Both the men's and women's clothing industries are almost completely organized and under union agreements. Textiles are now fairly generally produced under union contracts in the northern part of the country, and the unions have made some gains in the last stronghold of antiunionism, the South. Maritime transportation, except for a few tanker companies, is unionized. Longshoring on the Pacific is mainly in the hands of the CIO; on the east coast the International Longshoremen's Association, an AF of L affiliate, mainly controls. Even employees in specialty and department stores and in offices have joined unions (although not in great numbers) and work under union-employer contract arrangements.

A union contract is usually divided into four parts:

1. A preamble, which states the purposes and the scope of the agreement. For example, a contract between a steel company and the

United Steelworkers of America declares that it is the "intent and purpose of the parties . . . that this Agreement will promote industrial and economic relationships between the Union and the Corporation."

2. Legislative sections, which are the heart of the union agreement. They constitute the working rules and cover wages, hours, overtime, and other conditions of employment. The detail with which specific problems are handled is determined by the industry, the relations between the parties, and the history of union-management relations. For example, the agreement between a New England machine-tool company and the International Association of Machinists covers hours of labor per day and week, holidays and payment for holidays, overtime rates, distribution of work, pay for reporting-in when no work is available, lunch periods, bases for vacations, seniority, minimum rates of pay, handling of grievances, union recognition, and arbitration of unsettled disputes.
3. Terms providing for a judicial system to settle differences which arise over interpretation of its terms.
4. Provisions which state terms and means of enforcing the contract.

Most of the subjects covered by union agreements are self-explanatory. Some discussion is warranted, however, on several topics involved in agreements.

Seniority refers to the right of the worker to relative preference if he has been employed longer by the company than other employees. If, because of shortage of work for all of its employees, the company is forced to lay off some members of the working force, those who have had the shortest service with the company are laid off first. Seniority is also followed in rehiring, the workers with the longest seniority being the first rehired. In some cases seniority determines not only layoffs and rehiring, but promotions in the ranks of nonsupervisory labor. Thus, on the railroads a locomotive fireman's promotion to engineer is governed by seniority, as is promotion from trainman to conductor. Under some seniority clauses, a worker can claim the best job that his seniority will allow him to hold. Preference is based on seniority, and a worker with higher seniority can oust or "bump" a worker of lower seniority.

Some seniority clauses include a provision that seniority will apply when skill and efficiency are equal, but a rigid insistence upon that provision by management is likely to lead to difficulties unless differences in skill are readily measurable or are obvious.

Seniority rules were first applied on the railroads, but in recent years they have been widely introduced into manufacturing industry. Seniority begins on the first day of employment and continues throughout the duration of the employee's services. It has been maintained by some critics that seniority is inefficient because it places the burden of layoffs upon the younger workers. These charges may be true, but its defenders argue that seniority also serves to eliminate favoritism, protects the older worker who is not so mobile and might not be so easily hired by another employer, and sets up a labor pool upon which the employer can draw experienced help when needed. An employee who has been laid off will retain his seniority for a given period of time, but he must return within a limited number of hours when called back to work.

Recognition refers to the status of the union in a plant. A union may be recognized as the representative for its members only, or it may bargain for all workers, union and nonunion, in the shop. The *open union shop* is one in which union membership cannot be made a condition of employment but the union is recognized as the bargaining agent. A *closed shop* requires all employees covered by the contract to be members of the union, ordinarily at the time of hiring. Some contracts require the worker to join the union when he has been hired, while others require joining the union upon the expiration of a given number of days of employment. A *preferential union shop*—no longer a popular form—is one in which union men enjoy preference in hiring (and are the last to be laid off) but nonunion workers can be engaged if the union is not able to supply competent help. During the period of 1940-45, the pressure for a closed shop by the unions and the opposition of employers led to the development of a compromise—the *maintenance of membership requirement.* Under this contract provision, union members or those who join during the life of the contract must, as a condition of employment, continue their membership in the union. Under directives of the National War Labor Board, workers were allowed a 15-day escape period in which they could signify in writing their desire to withdraw from the union. A maintenance of membership or a closed shop clause is frequently accompanied by a *checkoff* of union dues. A checkoff requires the employer to deduct union dues from the pay of the employee and to remit these amounts to the union. The checkoff was first established in the coal-mine areas, and in recent years has spread to other industries.

Union-management contracts usually set up judicial procedures for the review of differences that arise during the term of the agreement. Differences in interpretation of specific clauses are almost inevitable, considering

the wide variety of topics covered by the agreement. An employee may be dismissed, for example, and the claim may be made that dismissal is too severe a penalty for the infraction involved. The agreement may provide that in the first instance the union steward—a representative of the union employed by the company (also called a union committeeman)—should attempt to settle the differences. If no agreement is reached, the contract may allow for one or more appeals, and finally for the submission of the dispute to a neutral party, called an arbitrator.

Some union contracts provide penalties for violations by employees. At one time an unauthorized strike in bituminous coal mining was punishable by a fine of $1 per worker per day. Violation of contract by union members is sometimes covered by "management security" clauses which provide penalties against the offending worker, such as denial of employment rights, loss of bonuses, or diminution of vacation pay.

GOVERNMENT AND TRADE-UNION ACTIVITY

Government attitudes toward trade unions have evolved over a period of 150 years from open hostility to protection and even promotion. The judicial doctrine which dominated the courts in the early cases dealing with labor was originally enunciated in the English courts. The main principle of this doctrine was that a group acting together has the power to inflict injury not possessed by the single individual. Thus, if a group of persons combine to accomplish an unlawful act, their mere combination is in itself unlawful and a conspiracy, although combination for a lawful purpose is not illegal. In cases involving social and economic questions, of course, the notion of what constitutes a proper purpose is influenced by social and economic predilections. American labor unions were first organized in an intellectual and moral environment of individualism which stressed self-help and *laissez faire.* The conspiracy doctrine was the basis for the prosecution of the Philadelphia Cordwainers in 1806, in which the defendants were in fact charged with conspiring to raise wages. In the course of time, some modification of this doctrine took place when the test became whether the *purpose* of combination was to inflict injury upon another.

The opinion of Chief Justice Shaw in *Commonwealth* v. *Hunt* (1842) marked an important turning point in the attitude of the courts. The opinion defined a conspiracy "as a combination of two or more persons, by some concerted action, to accomplish some criminal or unlawful purpose or to accomplish some purpose not in itself criminal or unlawful, by crimi-

nal or unlawful means." [3] In other words, the mere combination of workers could not be held as unlawful, but the test was the motives that gave rise to the action. This decision concluded that to raise wages by trade-union action unaccompanied by fraud, force, or intimidation was not illegal. The question of intent, therefore, became a guiding consideration. In other words, if the purpose is to injure the action is obviously illegal; if the injury is incidental to the object of benefiting labor, then the action is proper.

Injunctions

With the expansion of union activity in the 1880's and 1890's, unions were increasingly harassed by injunctions. The injunction is an order issued by a court of equity to restrain the performance of certain acts which, if committed, would inflict irreparable injury. A person whose property is damaged by another can sue to recover in the courts; but if the person inflicting the injury has no assets or the loss cannot be covered by a money payment, the remedy is obviously vain. The injunction is therefore a means whereby a person can prevent losses which are not recoverable through normal legal channels; in cases of labor disputes employers frequently appealed to the courts for injunctions against unions to restrain their actions.

Violations of an injunction can be punished by a citation and punishment for contempt, and thus union officers and their members can be severely restrained in many actions. Appeal to higher courts against injunctions is possible but also time-consuming and expensive. Frequently, in labor disputes the issues involved are settled and forgotten before the legality of an injunction is settled. Union activity is, moreover, seriously hampered during the period required to take the legal steps necessary. Unions became aware of the menacing nature of this instrument in the 1890's, and made efforts to curb the use of injunctions by the courts. Complaints were especially loud against blanket injunctions which restrained all persons, known and unknown. Not until 1932, with the passage of the Norris-LaGuardia Act, was the use of injunctions by the federal courts curbed. A number of states enacted legislation patterned after the federal act. The importance of the injunction procedure came to the fore in the coal strike during the winter of 1946-47 when it was used by the Federal government to restrain the president of the United Mine Workers' Union from voiding a contract covering government operation of the mines. The use of an injunction in this case by the government was upheld by the Supreme Court.

[3] Metcalf 111, 45 Massachusetts 111 (1842).

The Antitrust Laws

Not only were strikes subject to injunctions, but after 1890 the unions found themselves subjected to prosecution under the Sherman Antitrust Law. The Sherman law forbade combinations in restraint of trade or commerce among the states or with foreign nations; criminal and civil penalties could be imposed, and those injured by prohibited action could sue to recover triple damages plus costs. The law was directed against industrial combinations, but it was also applied to labor unions by the courts.

In the Danbury Hatters' case,[4] the United States Supreme Court held the hatters' union and its members guilty of violating the antitrust law. In an attempt to impose the union shop, a strike was called and boycott of Loewe hats followed. Sales fell and the company claimed in bringing suit that it lost $80,000 as a result of the boycott. Individual members of the local were liable for the damages, and many would have suffered heavy financial losses had not the American Federation of Labor raised most of the funds to pay the damages awarded. In the Buck's Stove and Range Company case, the American Federation of Labor and its affiliates were enjoined from conducting a boycott against the products of the company. Disobedience of the order led to the prosecution of three leaders of the federation. The United States Supreme Court upheld the right of the courts to prohibit an unlawful boycott.

Partly as a consequence of labor's opposition to the power of the courts in the issuance of injunctions, the Clayton Act (1914) included section 6, which declared "that the labor of a human being is not a commodity or an article of commerce"; antitrust laws were therefore held to be inapplicable to the legitimate activities of labor. But unions were not freed of responsibility by this legislation. A series of court decisions in the 1920's upheld severe limitation of picketing and boycotts by the courts.

Norris-LaGuardia Act, 1932

The labor movement redoubled its efforts to limit the use of injunctions in labor disputes and in 1932 Congress enacted the Norris-LaGuardia Act. The law regulates the issuance of injunctions in labor disputes. It requires that certain conditions must be met before labor activity is enjoined; it broadens the meaning of a labor dispute so as to include secondary boycotts, and it curbs the power of the judges to punish for contempt in event

[4] *Loewe* v. *Lawlor*, 208 U.S. 274 (1908); *Loewe* v. *Lawlor*, 235 U.S. 522 (1915).

of violation. Restraining orders can be issued only in the event that testimony shows an irreparable injury to the complainant's property will take place; moreover, the order is limited to five days. The complainant must have complied with all laws and show that he has made reasonable efforts to settle the dispute. Injunctions can be issued only after testimony in open court, implemented by testimony and cross-examination of witnesses. Those charged with contempt not committed in the presence of the court can ask for a change of venue and a jury trial. A number of specific acts are listed which cannot be enjoined as an illegal combination or conspiracy.

In a number of cases involving picketing, the United States Supreme Court has upheld the rights guaranteed by the Norris-LaGuardia Act. At present, the power of the federal courts to issue injunctions to private individuals or firms is sharply limited, and the states cannot outlaw picketing by blanket prohibition. They can, however, prohibit picketing in a context of violence.

The Right to Organize

Workers in their efforts to organize unions have confronted a variety of measures used by employers to prevent the formation and continuance of the unions. Among the obvious methods is the *lockout*—the refusal to keep members of a union employed. The employer refuses to deal with the union, and shuts down his plant: the lockout is a suspension of work initiated by employers. Another weapon of the employer is the *black list*, whereby one or more employers refuse to hire an otherwise competent worker because he is a member of a union. Some employers have not only black-listed union workers, but also required employees to sign an agreement not to join a labor union during their employment with the company. This is called a yellow-dog contract; and in *Hitchman Coal Company* v. *Mitchell*,[5] the United States Supreme Court forbade the union from attempting to organize workers who had signed yellow-dog contracts. The Court held that the yellow-dog contract is entitled to protection the same as any other legal right.

The Hitchman decision extended protection to the attempts of employers to insulate their workers from the union organizer. In an earlier case, *Adair* v. *United States*[6] the United States Supreme Court invalidated section 10 of the Erdman Act regulating interstate commerce, enacted in 1898, which forbade the discharge of an employee of an interstate carrier because of his membership in a union. The Court held that it was within the power

[5] 25 U.S. 229 (1917). [6] 208 U.S. 161 (1908).

of a carrier to discharge workers for any reason whatsoever, including membership in a union. Later in *Coppage* v. *Kansas*,[7] the United States Supreme Court invalidated a Kansas statute outlawing the yellow-dog contract on the grounds that the statute violated freedom of contract.

While these decisions had an adverse effect upon the position of organized labor, a countercurrent was developing at the same time. During World War I, the War Labor Board recognized the right of labor to organize and to bargain collectively through representatives of its choosing without coercion by employers. These principles were not entirely forgotten even in the antiunion mood of the 1920's.

An important milestone in labor legislation was the enactment of the Railway Labor Act in 1926. This law forbade coercion, interference, or use of undue influence by either party over the self-organization of the other, and furthermore established a means of peacefully settling disputes. In a unanimous decision the United States Supreme Court upheld the constitutionality [8] of this law. The practical result of the law was that an interstate carrier could not interfere with the organization and management of a railway union.

A more important step was taken in the direction of protecting the right of labor to organize by the inclusion of section 7a in the National Industrial Recovery Act. This clause guaranteed labor the right to organize through representatives of its own choosing, and forbade the requirement that an employee shall as a condition of employment join a company union, or abstain from supporting or joining a union of his own choosing. (A company union is one financed, dominated, or controlled by the employer.) With the invalidation of the National Industrial Recovery Act in 1935, the power to enforce section 7a ceased.

The National Labor Relations Act

Under the New Deal administration organized labor continued to press for statutory protection of the right to organize. In 1935, the National Labor Relations Act was adopted. This law is applicable to all industries engaged in interstate commerce. Section 7 provides:

> Employees shall have the right to self-organization, to form, join or assist labor organizations, to bargain collectively through representatives of their own choos-

[7] 236 U.S. 1 (1915).

[8] *Brotherhood of Railway and Steamship Clerks* v. *Texas and New Orleans Railway Company*, 281 U.S. 548 (1930).

ing, and to engage in concerted activities, for the purpose of collective bargaining or other mutual aid of protection.

Section 8 of the law prohibits employers from interfering, restraining, or coercing employees in the exercise of their right to organize, join, or form labor organizations. Employers cannot dominate or interfere with the organization or the administration of a labor union, nor are they permitted to support a union organization financially. Employers cannot encourage or discourage membership by hiring or discharging, nor can they refuse to bargain with representatives of their employees. The law was contested in the courts, and the issue resolved itself into an interpretation of whether manufacturing industries were in "interstate commerce," for the National Labor Relations Act was applicable only in industries engaged in interstate commerce. In a decision of the Supreme Court the applicability of the law was upheld. "Although activities may be intrastate in character when separately considered, if they have such a close and substantial relation to interstate commerce that their control is essential or appropriate to protect that commerce from burdens and obstructions, Congress cannot be denied the power to exercise that control." [9] In a series of cases supported by the courts the act has been applied to small as well as to large plants, to utility plants, to newspapers, and to mining operations. While agricultural workers have been excluded, the act has nevertheless been applied to citrus fruit packing and to lettuce packing.

The act created the National Labor Relations Board to administer its provisions. Complaints against employers' unfair labor practices are submitted to regional offices of the Board. If a basis for a complaint is found, a trial examiner holds a hearing, at which both the employer and the union can submit evidence. On the basis of the report submitted, the National Labor Relations Board may dismiss the complaint for lack of evidence, or order affirmative action, such as disestablishment of an employer-dominated union, or re-employment with back pay of employees discharged for union membership or activity. In the event the employer fails to obey the order, the Board can appeal to the United States Circuit Court of Appeals for enforcement.

In addition to preventing unfair labor practices, the National Labor Relations Board is authorized to conduct elections to determine bargaining representatives. Whenever elections are held, the Board must determine whether the bargaining unit be a craft, department, or industry. Frequently this decision on jurisdiction is extremely important, especially if two or

[9] *Jones and Laughlin Steel Corp.* v. *National Labor Relations Board*, 301 U.S. 1 (1937).

more unions are contesting the election. For example, if in a large steel mill a special bargaining unit is allowed for highly skilled patternmakers, it is likely that these workers might vote for a craft union. On the other hand, if a company or plant is the basis of the election, the skilled workers' votes will be overwhelmed by the relatively large number of votes of semi-skilled and unskilled workers. Problems of this kind have caused a great deal of contention among unions over the proper unit for bargaining purposes.

The Labor-Management Relations Act, 1947

A majority of spokesmen for industry attacked the National Labor Relations Act from its beginning. Amendment of the law was prevented until 1947, when Congress passed the Labor-Management Relations Act which changes in many respects the law of labor-management relations. Some of its provisions may be summarized.

This law prescribes legitimate rights of both employers and employees, and describes and restricts certain practices of labor unions and employers which have led to unrest. The new act modifies the Act of 1935 by declaring a number of actions of unions to be "unfair practices"; the original act was silent on this subject. These actions are: coercing individual employees to join a union; seeking the discharge of a worker in a union shop except for nonpayment of dues, initiation fees, and assessments; refusal of a union to bargain with the employer; and charging unreasonable initiation fees or dues.

Unions or employers must give 60 days' notice before terminating a contract or before demanding changes in the terms of the contract. They must also give a 30-day notification to the Federal Mediation and Conciliation Service before a contract termination. Existing contracts must remain in force during a 60-day period after such notice is given. Workers engaging in a strike during this 60-day "cooling-off period" would lose their rights under the law, but would regain them upon being rehired. The act also provides that either party to an agreement may bring suit for damages against the other for breaching the agreement.

In the event of a threat of a national strike in an industry vital to health and safety, the President may force a postponement of the strike for 80 days by injunction, during which time an emergency commission, appointed by the President, would investigate and report on the issues in dispute.

A union seeking certification from the National Labor Relations Board as a bargaining agent or requesting investigation of unfair labor practices

must file with the United States Secretary of Labor its constitution and bylaws, submit an annual financial report showing salaries of the three principal officers and all other officers whose aggregate compensation in the preceding year exceeded $5,000. The report must also show the dues and initiation fees charged; the way in which officers are chosen; any restrictions upon membership; the manner in which strikes are voted; and the manner in which money is spent. These reporting provisions are aimed at undemocratic practices of unions and domination of certain unions by their officers.

Officers of unions seeking protection of the law must certify in writing that they are not members of the Communist party and do not believe in the violent overthrow of the government of the United States. This provision is designed to weaken the position of communists in the few unions under communist control.

Union welfare funds supported by employer contributions must specify the benefits to be given to the employee, must be annually audited, and must be jointly managed by employer and employee. Checkoff or deduction of dues by an employer from the pay of an employee is allowed only when authorized by the employee in writing. Such authorization can also be revoked by the employee. Unions are not permitted to donate money for use in any national election, primary, or convention.

This legislation is regarded by union leaders as a reversal of national policy in protecting the position of unions. By employers generally it is looked upon as a correction of the "one-sidedness" of the original National Labor Relations Act. The effects of this 1947 legislation upon the position of unions, union-management relations, and union relations with their members cannot be judged at the time of writing. Union opposition and violation of the law will lead to a test of the constitutionality of many of its provisions; union political activity may lead to its change; and union-management contracts may in effect nullify some of the provisions. Whatever its eventual development may be, the Labor-Management Relations Act of 1947 represents in part a reaction against the growth in power of unions and in this respect is similar to legislative attempts to curb private power in industry.

PEACEFUL SETTLEMENT OF LABOR DISPUTES

Differences arising over terms of employment may be settled peacefully, i.e., without resort to economic pressure by strikes and lockouts. Most differences are in fact settled in this manner, for at any one time there is

usually not more than a small fraction of the working force engaged in strikes.

Conciliation

Conciliation and *mediation* are terms used interchangeably for the process whereby an outsider—usually a government officer—aids in the settlement of differences between a union and an employer by participating in discussions. The conciliator is usually an experienced person whose presence is intended to clarify issues and prevent precipitous and perhaps unwise ending of negotiations by either party. Mediation is usually carried on by state and municipal officials on an *ad hoc* basis, although some states have specific boards for this purpose. Usually the conciliating officer is from the state Department of Labor or the Federal Mediation and Conciliation Service, which was formerly a branch of the United States Department of Labor and since 1947 is an independent agency.

Voluntary Arbitration

Another means of settling disputes is through arbitration, which consists of submitting the dispute to one or more outsiders who give a decision after hearing the evidence of both parties. Usually the parties are obligated to accept the decision. An arbitration board may be made up of a representative from each side, with one or more neutrals; or all of the members may be neutral. Arbitration may arise under two circumstances. A large majority of union contracts provide that during the term of the contract disputes that cannot be settled by negotiation will be submitted to arbitration. In a number of industries permanent arbitrators have been appointed, sometimes called impartial chairmen, to whom all unsettled disputes are submitted. Arbitration may also be used to settle disputes arising during negotiation of a new contract or renewal of a previous one. During the term of a contract both parties are usually required to submit unsettled issues to arbitration, but there is no such obligation when a new or renewed contract is being made.

Compulsory Settlement of Labor Disputes

Labor disputes may be settled by a government agency. In Canada compulsory investigation was established first in industries defined as public utilities. Ten days' notice must be given by one party to the other if a

change in the conditions of employment is intended. If a dispute over the intended change develops, no strike or lockout may take place until a government-appointed fact-finding board has made its reports, after the submission of which a strike or lockout can be called. Basic to this arrangement is the belief that public pressure will force an acceptance of the fact-finding committee's report. The use of emergency boards to investigate and report on disputes that arise in the interstate railroads in the United States is another example of the fact-finding method which delays direct action by the disputing parties. The Labor-Management Relations Act of 1947 also provides for a fact-finding commission when the President uses his authority to obtain an injunction to delay a national strike imperiling public safety or health.

Compulsory Arbitration

Australia and New Zealand have experimented with compulsory arbitration. In both countries unions voluntarily register with a government agency. Since employers are reluctant to deal with nonregistered unions, and since nonregistered unions cannot participate in certain types of government activity such as voluntary arbitration and conciliation, unions have usually been eager to register. Disputes in New Zealand come before a Court of Arbitration which has authority to impose fines for violation of its decisions. At times employers have opposed the system, and at other times the unions, but it has worked reasonably well, especially under the Labor government. Labor problems are by no means suddenly solved by this method nor do all difficulties vanish. In most cases, there is a question whether the same ends cannot be attained by unhampered collective bargaining accompanied by a reasonable attitude on both sides, because a system of compulsory arbitration must be accompanied by such an attitude if it is to be successful.

SOME LARGER ISSUES

The growth of large unions such as those in the steel, automobile, railroad, communications, and other major industries, together with the concentration of business control in the hands of a relatively small number of large corporations, raises some fundamental issues in American economic life. For what purposes will this economic power be used? The union official and the corporate executive are representatives of private group interests, although both may profess a larger social interest. When the coal

industry and other large-scale industries like the steel, automobile, and railroad industries, are hampered in their operation by disputes between employers and unions, the entire nation is affected. Both large companies and large unions, which are increasingly centralized in the hands of a small number of men, represent the private interest of respective great monopoly power. Both groups may affect prices—of labor and products—and both are at times in a position either to promote peace and a co-operative spirit or to disrupt the national productive process. The rise of "big government" has been in part necessary as a controlling force over the powerful agencies of concentrated industry and unionism.

It is important, therefore, to regard the growth of unionism and the concentration of corporate control in terms of (1) the centralization of economic power, and (2) the purposes underlying the exercise of that power. Before concentration and centralization were so dominant in industry and unionism, the disputes between unions and management were frequent and often violent, and always costly to the communities involved. However, these disputes usually were localized. As both parties grow in size their disputes become increasingly nationwide in scope. The coal and railroad trainmen strikes of 1946 are dramatic instances of this phenomenon. Similarly important are the disturbances that occur in relatively small industries producing essential parts for many other industries; labor disputes in the few firms supplying such parts can cause the disruption of production over a wide area. Thus union officials and the executives of the few firms have in a sense the responsibility for continuance of productive activities in the nation.

What are the issues in wage and working-condition disputes? From the corporate manager's point of view, granting a few additional cents an hour might raise production costs and reduce the profitability of the firm to a point of financial difficulty. The executive might also have an eye to the future. Although present business activity may warrant the wage rise, business might fall off a year hence and the increased wage costs might cause losses. In addition to these business considerations, the executive might fear that a wage concession will be interpreted by the union as a sign of weakness and so encourage further wage demands in the near future.

The issue for management might be complicated by larger factors. If a firm is part of a large industrial empire, the other industrial interests and possibly the large bank or banks controlling the firm might look upon the dispute as a "test case" which will determine union demands upon other firms in the industrial empire. These outside industrial and banking interests might insist upon "holding the line" to prevent high wage demands

being made against other firms. Furthermore, the firm might be impelled to resist demands because of personal factors. Union leaders involved may be *personae non gratae;* the firm's resistance to their demands could be designed to embarrass them, and possibly to encourage a rival faction within the union. Or, if the corporate management is under fire from important stockholders, a concession to the union—even though financially possible—might unseat the management. Or the management of the firm might have deep-seated personal objections to collective bargaining and refuse further wage concessions "on principle."

Considerations from the union's point of view may be equally complex. The wage issue might be one that involves rectifying a number of "inequalities" in the wage scales in the firm, i.e., some workers are paid a wage which is unfair in relation to other workers' wages. The union insists upon the higher average wage rate as a means of eliminating such inequalities. Or the wages in the firm might be somewhat lower than those paid by other firms in the same industry and in the same communities. If the firm is earning substantial profits, the union may support its otherwise valid claim by a dubious "ability to pay" argument. The union attitude in the dispute is usually influenced by its past relationships with the firm. If the firm had traditionally opposed the union and its past demands, the union for this reason might be adamant in its current demands. The union, too, might regard the wage dispute as a "test" case. Should its wage negotiations be regarded by the national union as the opening skirmish in a nationwide campaign to raise wages, the local union will be urged by the national board and by other unions to insist upon their demands with the possible inducement of financial and other aid if a strike ensues. Personal and organizational considerations determine the attitude of the union officials in their negotiations as well as management. The position of the leaders might be threatened by a rival internal faction or by a rival union; success in negotiation may mean maintenance of control.

From the point of the consuming public, the importance of a wage rise by the firm will depend upon whether the firm passes all or part of the added wage cost to the selling price of its product. This depends in part upon the elasticity of demand for the firm's product. However, if the dispute is a "bellwether" case, the public's interest is much broader. The size of the wage increase and of the subsequent price increase, if any, will determine the extent to which the public's dollar depreciates in purchasing power. If all major industries follow the leader and raise both wages and prices, all who do not receive an increase in income corresponding to that received by the organized workers stand to lose. The position of the dis-

abled, retired, and others on fixed incomes is worsened more than that of wage earners.

The public's interest in strikes is even greater. Strikes in major industries are not simply private disputes. They throw workers in other industries out of jobs, disrupt production throughout the economy, and cut off or endanger services necessary to safety and health in a large part of the country.

The growth of unions and the concentration of corporate enterprise has essentially transformed the collective-bargaining process, since it imposes upon both parties far-reaching economic responsibilities. The "industry-wide collective bargaining agreement" is a means by which responsibilities are joined in large industries; in an economic sense, a monopoly of labor and a monopoly of enterprise decide—together—by negotiation and a contract the conditions of employment and the profitability of investment. In other industries, while a general contract is not executed, the terms decided by the union and a leading company set the pattern for contracts with other companies. The prices of the commodities produced, the profitability of investment in the industries concerned, wage rates in industry in general, the volume of output and income—all are influenced by the decisions thereby reached. Collective bargaining in the major industries as a consequence no longer remains confined to the private interests of the immediate participants.

From this realization comes the major question of the present day. Should these huge aggregates of economic power be broken up, or should their relationships be rigidly controlled by government, or is there any other alternative? The break-up of monopoly has been attempted with dubious success; and breaking up the big unions by law promises no better results, if indeed it is politically feasible. Limiting the scope of unions' activities, without affecting their size, is sometimes advocated. The abolition of the industry-wide agreement by law has been proposed; other legislative proposals seek to limit collective bargaining contracts to a regional basis and even to the individual plant basis. This might merely multiply the making of contracts with little other effect, because each party would tend to strive for uniformity in the terms which were advantageous to it.

Another proposal, namely, governmental intervention to assure acceptance of collective-bargaining terms, is in effect a negation of collective bargaining. At all points of dispute the decision of government would rule, and it is likely that the economic considerations in the wage bargain would be replaced by political considerations.

Fortunately, employers and employees get along for the most part. Dis-

putes over wages and other conditions of work normally affect only a small fraction of employed persons at any given time. The public interest is materially involved only when a vital productive service is disrupted, or when disputes range widely over the economy causing large numbers to be out of work, or when employers and unions raise both wages and prices to the disadvantage of the public at large. Government may, as it has attempted to do by the 1947 legislation, mitigate these difficulties somewhat by defining the rights and responsibilities of both parties to the transaction, and by making both parties equally responsible under law. Legislation, however, is no assurance against the abuse of power by either or both, for it may succeed only in establishing a third aggregate of power designed merely to establish the rules of behavior for the private monopoly power of unions and management. The issues arise fundamentally from the right of free association in this society. Differences of opinion and the attendant disputes are from a long perspective to be welcomed and not deplored, for they are symptoms of the right of free association. So long as these disputes remain *economic* in character the existence of a free society is not threatened; however, should the issues between management and labor become centered in opposing political camps, as they have in Europe, they would then become dangerous to the present form of American society.

SOME ECONOMIC CONSIDERATIONS

The wage- and salary-earning group accounts for about two thirds of the number of gainful workers. The income position of this group may be described on the basis of (1) its *share* in national income, and (2) its average earnings on an hourly, weekly, or annual basis.

During the prosperity of the late 1920's, wages and salaries comprised between 65 and 70 per cent of total national income; the share in 1946 was approximately the same. In contrast, this group received 80 per cent of the national income of 1932. Thus the *share* of the total going to labor is somewhat larger in a period of depression than in a period of relatively full employment. The actual *amount* of income received as wages and salaries, however, is less in depression than in prosperity. Another comparison of importance is the constancy of employees' share in the late 1920's and in the mid-1940's. The aggregate income in 1946 greatly exceeded that of 1929 but employees received about the same share of the total in both periods.

For the amounts earned per hour, week, or year, the figures show considerable changes. Available wage data are not adequate to permit a full

comparison over a long period of time but a general indication may be obtained from certain large segments of the employee group. From 1914 to 1946, the average weekly income of wage earners in manufacturing, railroading, and coal mining increased approximately 400 per cent. Over this period the cost of living approximately doubled; as a consequence the rise in the *real* wage was less than the rise in money wages. However, the rise in prices and the resultant increase in living costs in turn are offset in some measure by improvements in the quality of many things comprising cost of living items. Real wages, therefore, have increased substantially over the three-decade period.

From the late 1920's to 1946, the average annual earning of all wage and salary employees increased approximately 60 per cent. In the same period, average weekly wages in manufacturing increased 85 per cent and similar advances were made in the railroad and coal-mining industries. Wage rates per hour advanced somewhat more, but hours worked per week were less. Again, part of this increase in money income was offset by the somewhat higher living costs in 1946 compared with 1929, but the rise in real wages was nonetheless substantial.

The evidence at hand thus indicates that wage earners in recent years are considerably better off than in the prosperous 1920's, and much better off than in 1914 or during the 1930's, if well-being is to be measured in terms of annual, weekly, or hourly real wages.

The influences that determine the income and therefore the material well-being of employees may be placed in two categories: (1) those governing the level of national income, and (2) those governing the share or percentage of national income received by employees.

The conditions governing the size of national output, income, and employment have been discussed fully in Chapters 4 to 7. Of special importance is the volume of investment expenditure when this is related to the saving-spending ratios of the economy as a whole. The growth of national income and output over a period of time is possible, if the volume of investment increases, and if the propensities to invest and to save correspond. Investment in capital facilities adds to productive capacity; the progress of technology and the increase in the supply of labor in the course of time also augment the nation's productive capacity. The relatively full use of this productive capacity creates the real and money income which permits a rise in the material well-being of employees, and other groups as well.

The influence of unions on the level of national output is difficult to assay. Their *direct* influence is limited. To the extent that unions are able to control the quality and quantity of effort expended by their members while

at work, they can influence output per man as well as aggregate output. Some unions encourage increased output per hour or day on the part of their members; others are indifferent; still other unions actually limit output by restricting the amount of work done. The net effect on national output can hardly be measured. Beyond this, unions have no direct influence, for they are not important investors in the sense relevant here. Nor by their expenditures do they add to aggregate demand otherwise, for the funds they spend are drawn from their members' income and would probably be spent in any case.

It is conceivable that unions can exert an indirect influence on the level of national output and income by means of the spending-saving ratio. That is, if unions succeeded in securing for employees a larger share of total national income through wage increases, the spending ratio would tend to rise. Whether this would be conducive to future increases in output, income, and employment would depend upon the state of long-term profit expectations and the effect of wage advances on these expectations. If these profit expectations were not very optimistic, and if the wage increases did not make them still less optimistic, then the larger volume of consumer expenditure (because of wage increases) would bolster output. These are *possible* indirect influences favorable to expansion. It will be remembered that the mature economy thesis places considerable confidence on these indirect influences, while the expansive economy thesis takes a contrary view.[10]

Consider now the influences governing the share of total income received by employees. This share was, as previously stated, about the same in 1946 as in 1939 and 1929. The influences regulating this share include (1) wage rates, (2) the amount of capital employed in production, (3) and the nature of business organization.

An increase in wage income, not offset by (1) a rise in prices of goods produced, and (2) an increase in output per worker, would increase the share of total income going to employees. Unions succeed in obtaining higher wages for their members. In spite of sizeable wage increases, however, the share of national income received by employees was no greater in 1946 than in 1929. This need not mean that the efforts of unions to better the income position of wage earners availed nothing. Had unions not succeeded in securing wage increases of the magnitude attained, the share of employees in total income might have fallen. This is a moot point. The wage increases—comparing 1946 with 1929—have been balanced by a rise in other kinds of income and by the introduction of new and improved

[10] See Chap. 7, pp. 153-65.

technology, for wage and salary costs (income) have remained at approximately two thirds of national income. The relationships between wages, prices, and technological developments (to reduce labor costs) are therefore complex. A wage increase in some instances might induce a price rise, or it might lead to the introduction of laborsaving machinery. On the other hand, a wage increase might follow such changes partly because of union action and partly because of other labor market conditions. A combination of these relationships, in the years under discussion, undoubtedly operated to leave the share going to employees unchanged.

The effect of a growth in the amount of capital employed in the production of national output was suggested in the foregoing. If an increase in capital occurs with no concomitant increase in the demand for wage earners, the wage earners' share in total income will fall. In some instances unions have resisted the introduction of capital facilities that displace labor. In general, however, they have rather emphasized the payment of higher wages as production has increased with the growth of capital. Wages will tend to rise with the increase in capital in any case because, as capital resources grow, labor becomes relatively more scarce and its marginal product increases. The influence of unions tends to make such wage increases more quickly responsive to the growth of capital.

The share of total income going to wage and salary workers is affected by the character of business organization. The growth of monopoly might be expected to increase the nonlabor share in total income. As sellers of goods and as buyers of labor service monopolists can derive substantial advantages. Unions have no influence on the monopolists' sales strategy; however, by confronting the monopolistic buyer of labor service with a monopoly in the supply of labor service, unions are in a position to reduce the advantage of the former. Indeed, it is in the industries dominated by large corporations that unions in recent years have grown most rapidly. In these industries hourly wage rates tend to be higher, having increased relatively more, than in the less concentrated industries. To some extent unions are responsible for these wage advances, although the growth of capital as such in large industries had led to higher wages before unionization was effective.

To sum up, the amount of income that wage earners receive and the share of the total going to them both depend upon the size of the national income; and the share in turn is furthermore influenced by wage rates, the amount of capital, and the structure of business organization. Although unions have little to do directly—and possibly not much indirectly—with the forces that govern the size of the national output and income, they are in a position

to influence the other factors. They do so by virtue of control over given parts of the supply of labor service. Those workers who are embraced by union organizations are in a relatively better position to secure income advantages than unorganized workers as a whole. Wage rates of union workers are in general above unorganized workers in similar types of work. However, the advantages of unionism arise partly because of the limited coverage of unions. If all wage earners were organized some of the special advantages, including increased wages, of organization would disappear.

CHAPTER 30

Government and the Economic System

Governmental participation in the economic affairs of the nation has grown and changed significantly in the course of this country's history. At the present time, the federal, state, and local governments exert an important and pervasive influence, partly because of their own activities (such as national defense and public education) and because of their regulation of, or influence upon, the decisions of other economic agents. To weigh the role of government in the economic life of the nation is indeed difficult. Governmental expenditures as a component of aggregate demand have ranged in recent years from 15 to nearly 50 per cent of total expenditures for the commodities and services composing the national output. But the influence of government is by no means limited to its share in the aggregate demand for national output. By its innumerable regulations of business behavior, tax policy, labor legislation, and other forms of participation, government affects the decisions of other sectors of the economy, both in their expenditures and in their use or allocation of economic resources.

The economies of western Europe, Great Britain, Canada, and the United States in the nineteenth century were characterized by a *laissez-faire* policy and a philosophy of economic liberalism. The role of

government in economic affairs was relatively slight, for in the spirit of the age an exaggerated individualism was foremost. But an economic system founded on individual private enterprise, with markets and prices as the mechanism to co-ordinate the decisions of individuals, cannot exist as if in a vacuum. Such a system requires a framework of political stability, of orderly government, such as established for this country by the Constitution. Some amount of law and order is required so that the production, exchange, distribution, and consumption of economic goods and services can proceed at all. But even when, as in the nineteenth century, the functions of government were at a minimum, the role of government was not a passive one. Government provided aids of both a general and a particular type to the economy, to encourage its growth and development. The provision of a standard system of weights and measures, the construction of postal roads, the establishment of a patent system giving monopoly rights to inventors, together with the formation of a money system and the subsidization of manufacturing through tariff duties, were among the first "controls" over private enterprise. These controls were clearly promotional rather than restrictive. In the course of time the growth of controls—to promote business activity, to regulate this activity, to restrict certain types deemed undesirable, and to protect industries or groups—has proceeded far, so that at present the American economic system is far removed from the economic liberalism of the last century. Economic and social problems, war and other national emergencies, have called forth the extensive participation of government in spheres once considered the exclusive domain of the individual and of the impersonal operations of the market.

The role of government—whether it be highly restricted or all-pervasive—is defined and sanctioned by law, by judicial interpretation, and by the sentiments and views of the politically dominant portion of the public.

CONSTITUTIONAL PROVISIONS

The Constitution broadly defines the spheres in which government may actively intervene as well as the rights of the individual in the conduct of his economic activities. This document is by no means solely or even largely an economic blueprint, but it does establish basic legal precepts relating to the economic role of government. In the course of time, both through amendment and changing judicial interpretations, this role has broadened enormously. The Constitution's economic provisions may be traced from the economic and social problems confronting the nation in the immediate post-Revolutionary war years. With the close of the war Euro-

pean goods flooded the American market, threatening the small industries of the period; at the same time European countries discriminated against American imports. These problems were intensified by interstate tariffs, as each state sought to foster industry or raise revenue at the expense of others; state governments depreciated their currencies, passed laws impairing creditors and favoring debtors, and otherwise threatened the position of the propertied groups. The Constitution in its economic aspects sought to protect the new economy from disruptive foreign policies, from the divisive actions of the separate states in the Confederation, and from inroads upon personal and property rights by "irresponsible" groups. The powers granted to the Federal government were limited but nevertheless adequate to guard against these disruptive influences.

Of great importance is the power to regulate commerce with foreign countries and among the states. The "commerce clause" grants to the Federal government jurisdiction over interstate commerce, and in the course of time became the constitutional basis for the regulation of many forms of economic activity; indeed, the first major control instituted by the Federal government over business—railroad regulation—was based on this constitutional power. In recent decades Supreme Court interpretations have greatly extended the jurisdiction of the Federal government to economic activities formerly regarded as beyond its power; for example, in the Jones-Laughlin decision upholding the National Labor Relations Act, the Court held that labor relations in mass production industries directly *affect* interstate commerce, i.e., a strike would stop interstate commerce. This interpretation brings within the orbit of federal control any industry whose activities may affect interstate commerce, or any problem which may be construed by Congress and the Supreme Court to affect interstate commerce.

Both the Fifth and Fourteenth Amendments to the Constitution have decisively influenced the economic life of the nation. The Fifth Amendment provides that "no person shall . . . be deprived of life, liberty, or property, without due process of law . . . ," the Fourteenth "Nor shall any State deprive any person of life, liberty, or property without due process of law. . . ." The effect of the "due process clause" has been to safeguard private property and private enterprise from governmental encroachment. Judicial interpretation of these constitutional amendments has also varied considerably. In 1895 the Supreme Court held an Illinois eight-hour law for women workers unconstitutional because it interfered with the (property) right of a person to dispose of his labor freely. Since then the interpretation has been liberalized so that social legislation affecting hours and working

conditions of special classes of labor is constitutional, i.e., the Court holds such state legislation to be a proper exercise of police power to assure public safety. The due process clause has often been invoked to invalidate restrictions placed on business by both the Federal and state governments. Government may and does abridge property rights when abridgement is deemed necessary for the public welfare or when it occurs as a consequence of clearly defined powers of government; at all times, however, governmental encroachment is held within limits defined by constitutional provision as interpreted by the judiciary.

The Constitution grants residual powers to the states, i.e., powers not specifically granted to the Federal government are reserved for the state governments. Their powers, therefore, are extensive. Intervention in economic activity by the states may be an exercise of "police power" designed to protect the health and welfare of their citizens. In general, regulation of business activity by state governments (e.g., child labor, factory conditions, labor union practices, public-utility rates) is subject to the condition that public benefits exceed the private sacrifices imposed by law. The judiciary occupies the important position of assessing the respective benefits and sacrifices, thus limiting the power of legislatures in regulating business activity.

This discussion merely touches on certain aspects of law which define the power—and limit the power—of government to intervene in economic activity. The Constitution imposed a large degree of uniformity on the legal structure of the country; it guaranteed a large measure of freedom to the individual to pursue his own ends; and it safeguarded property rights. The scope of governmental powers, of individual freedom, and the guarantee of property rights are by no means rigidly defined. The courts interpret these provisions as issues arise; their interpretations have gradually extended the permissible intervention of government and have restricted the permissible actions of individuals and the rights of property. These changes in interpretation often accompany (or follow) changes in the character of economic life. For example, regulation of manufacturing industry was for long considered outside the jurisdiction of the Federal government; both in a legal and economic sense it was "local" activity. The growth of large-scale industry removed many segments of manufacturing from the local category in an economic sense; markets for raw materials and for the finished products became regional or nationwide. In the course of time both Congress and the judiciary accepted this fact; since such industry *affects* interstate commerce it comes within the orbit of (limited) federal regulation.

Legislative enactments and judicial interpretations occur not in a vacuum

but rather in a context of economic and social problems and public views and attitudes. The abridgement of property rights and of the freedom of the individual in his economic activities has received the general approbation of the public in recent times. The "climate of opinion" today allows a degree of governmental intervention which would have been repugnant to the highly individualistic views prevailing in an earlier period. In part, the complexity of the economy, the great interdependence of its parts, and especially the urbanization of the modern period impel an increased reliance on government.

The discussion turns now to the principal areas of governmental control, to some of the problems encountered in regulation, and to governmental ownership. It is too much to expect clear-cut policy on governmental intervention, for much of it arises from the varying political pressures of interested groups and from the emergencies of war and depression. Frequently the original purpose of control undergoes modification and indeed may be perverted to serve the ends of special interests. In a predominantly free-enterprise system such as that of the United States, intervention is of necessity piecemeal. It is intended to contribute to the "general welfare" more than it detracts from those whose activities are curbed or regulated. This purpose is sometimes a disguise for what is actually private interest, and, at best, the measurement of general advantage and private disadvantage is difficult.

AREAS OF GOVERNMENT CONTROL

The controls over economic activity exercised by government are of three main types: (1) those which benefit private enterprises in general, (2) those which benefit particular industries or groups and which also confer general advantages, and (3) those which assist particular industries or groups and either confer no general advantages or create some disadvantages. Of the first type, the various measures preventing fraud are examples. Subsidies to transportation are an example of the second type of control. The transport companies are benefited, but so is everyone who gains an advantage through lower costs of transportation. (The taxation required to raise the funds paid out as subsidy is in general the measure of the disadvantage entailed by such controls.) The third type of control is exemplified by tariff protection to an industry without significance in the national defense. One industry gains at the expense of others, and there is an additional burden upon taxpayers or consumers.

The Regulation of Markets

Markets are economic areas in which buyers and sellers meet, determine prices and the other terms of exchange, and thus affect the allocation of economic resources. There are tens of thousands of markets, because of the large number of commodities and services and because of the many (physically separate) areas in which they are bought and sold. Most markets are not subject to public administrative regulation; they are regulated by the customs and practices of the buyers and sellers in them and by the influence of law. Some markets, however, present economic and political problems which have led to their being controlled in one way or another by government. This is obvious in the case of labor markets, where collective bargaining is affected by the operation of the labor relations legislation. Where monopoly necessarily exists, as in the sale of public-utility services, some form of public administrative or governmental regulation is present. The organized commodity markets in farm products are also subject to some government controls, because the prices arrived at in these markets are all-important in determining farmers' incomes, which have become a special concern of the general economy through the action of the strong farm bloc in Congress. Furthermore, the securities markets have been put under government regulation.

The regulation of the securities markets, i.e., the stock exchanges, and of the issuance of new corporation securities affects virtually every corporation and the millions of owners of corporate securities, and indirectly the much larger group of insurance policyholders and of savings-bank depositors. The flow of investment and the changes in business profit prospects are vital in the functioning of the economy, and these are in part determined by the issuance of securities and the fluctuations in the stock markets. The purpose of governmental regulation of the securities markets is the comparatively modest one of preventing fraud; it is based on the belief that complete and honest information will enable the buyer to make wise choices. This regulatory activity has been discussed in Chapter 26.

It is important to realize that prices in the stock markets, except in the most indirect way, are not regulated. What regulation exists is, in a fundamental sense, superficial, because it amounts at most only to reduction of dishonest and underhanded practices in buying and selling securities. Price formation is still free and is a resultant of individual profit-seeking decisions. The regulation of new security issues affects neither the total volume of new investment nor the direction of investment, except to the slight extent

by which volume and direction can be influenced by the availability of better information.

The commodity exchanges, in which wheat, corn, cotton, and other staple products are bought and sold, function very much as do the stock exchanges. The collapse of the boom of 1918-20 brought with it a very sharp fall in farm prices. The subsequent complaints of farm organizations blamed the markets themselves as well as the speculators in these markets. In legislation passed in 1921 and 1922, Congress provided for some slight regulation of these markets by the Department of Agriculture. In 1936 Congress extended and intensified these controls. While the principles are the same as those of stock-market regulation—the furnishing of information and the prohibition of specific manipulative practices—commodity exchange regulations have gone farther. For example, there are restrictions on the size of speculative operations by any one trader at any one time, and there are limits to the extent of price change per day for some of the commodities. Such regulation influences only the slight day-to-day movements of prices. The longer term and therefore more important price variations remain unaffected.

Maintenance of Competition

The antitrust laws have already been described in connection with business organization and on the concentration of industry.[1] In the following brief discussion the antitrust laws and the broad public policy which they represent are described in the context of government control as a whole. Essentially, the purpose of the Sherman Act of 1890 and of the Clayton Act of 1914 is to maintain competition as the regulative force of private enterprise. In a sense, Congress has delegated to competition the tasks of controlling prices and of determining the allocations of resources among industries. But as was shown in the earlier chapters, the antitrust policy has not been entirely successful in inhibiting monopolistic tendencies in the economy. This has been partly a result of restrictive court interpretation, partly a result of lack of zeal and funds for enforcement, and also a consequence of the oversimplified assumption concerning the nature of monopoly and of competition which underlies the antitrust policy. The traditional view is that in any industry or in any market there exists either competition or monopoly; one is good and the other is bad. Actually there are many forms of competition, and some of them, such as oligopoly, may lead to a resource allocation not much different from that which would occur under monopoly. Since the antitrust laws at best can accomplish only the elimination of monopoly and

[1] Chap. 27.

of monopolistic practices, the reliance upon competition does not afford the certainty that resources will everywhere be allocated according to costs and to consumer incomes and preferences.

Another aspect of the maintenance of competition through government controls is the regulation of methods of competition. The complexities of competitive methods were discussed in Chapter 16. There are dozens of separate competitive practices which have been declared unfair under the operation of the Federal Trade Commission Act of 1914. This act and its amendments are administered by the Federal Trade Commission, whose duties are, stated simply, the policing of monopolistic competition. Business firms which resort to the unfair practices which are unlawful are ordered by the FTC to cease and desist. Examples of such practices are false or misleading advertising, misbranding, selling rebuilt articles as new, and misrepresentation.

The Robinson-Patman Act of 1936 is also administered by the FTC. The act is aimed primarily at chain stores, and has the purpose of improving the competitive position of the independent retailer and of the wholesaler by prohibiting price discrimination by manufacturers in selling to their different classes of buyers. Manufacturers had previously given price concessions to chain stores which often went beyond the extent justifiable on the basis of quantity discounts. In principle, the Robinson-Patman Act requires manufacturers to sell at the same prices to all types of buyers, whether chain stores or independent retailers. Quantity discounts are still permitted but must be based on cost differences. Chain stores are thus permitted to utilize whatever cost advantages their size gives them, but are not allowed to use their size as a bargaining factor.

Although the maintenance of competition is the objective of the government controls which have just been reviewed, other forms of regulation have the contrary effect. *Resale price-maintenance* laws tend to limit competition by restricting the freedom of retailers to vary the prices they charge. These laws, also called fair-trade laws, have been adopted in one form or another by nearly all the states. In passing the Miller-Tydings Act of 1937, Congress removed the possibility of conflict between these state laws and the federal antitrust laws, and thus in effect gave approval to the resale price-maintenance laws, although the FTC had held that resale price maintenance is a method of unfair competition. Under the state laws now in force, manufacturers may themselves fix the retail prices of their products, and the retailer is required to sell at these prices. This type of legislation was adopted in the states as a result of the pressure of independent retailers (and some manufacturers), who sought to mitigate the effects of competition

from chain stores and other retailers following low-price policies. The legislation is in the interests of some retailers and of some manufacturers, it harms others and is on the whole detrimental to consumers. Chain stores and other large retailers have, however, largely freed themselves from the restrictions of resale price-maintenance laws by introducing additional brands of their own.

In other ways also the states have limited competition. "Interstate trade barriers" had become a problem in the period just before World War II. The barriers are a series of indirect controls which limit the movements of commodities across state borders and which restrict competition from out-of-state business firms. In nearly all cases, the barriers have some other ostensible purpose than the limitation of competition, for interference with interstate commerce per se by the states is unconstitutional. For example, the protection of consumer health in milk inspection regulations often has, and not by accident, the effect of preventing milk from being shipped from one part of the country to another where there is demand for it.

The sale of foods, drugs, and of woolen products is the object of special federal control, which has to do primarily with accurate labeling of these products. Meat is Federally inspected and graded. The states also engage in such regulation, with particular attention to such commodities as milk. It would be a mistake to suppose that this type of regulation is primarily intended to benefit the consumer. Consumer advantage is almost always a by-product—though an important one—of protection to the producer. Take wool-labeling as an example. The consumer is protected from deception and finds every article or garment accurately labeled. But the purpose of the legislation, and the cause of its being adopted, was to protect wool-growers from competition from reworked wool and from other materials. The consumer interest is rarely organized as effectively as the producer interest.

Public-Utility Regulation

The term *public utility* will be used here to embrace all industries which are deemed by law to be vested with a public interest and which are therefore subject to special regulation by government. They are for the most part private enterprises which occupy a special position in the economy. The public utility is peculiarly an American institution, for the industries which in this country are conducted as public utilities have for some time been publicly owned and operated in Europe.

It is not possible to state with precision and in a way which would

command general agreement just what "public interest" is. Whenever the legislatures and courts agree that a certain business is vested with a public interest, then that business is subject to some form of regulation as a public utility. Usually, although not always, the element of monopoly is present and the product or service is considered indispensable. Electrical energy is a leading example: geographical monopoly in production, and indispensability for most consumers in the present state of technology, are characteristic. On the other hand, passenger bus service is not always conducted by monopolies, but public transportation, including passenger bus service, has been considered of such special importance that transport companies have long been public utilities. In air transportation there is neither monopoly nor indispensability of service; yet for other reasons—safety, national defense, and speed of communication—this industry is also a public utility. The list of public-utility industries is a long one. It includes transportation (rail, water, motor carrier, and air), communication (telephone, telegraph, and radio), some forms of energy (electricity and gas), and a series of such industries as milk distribution, warehouses, and ice plants. The concept of a public utility differs among the states, and has changed over time, the tendency having been for the list of public-utility industries to be expanded.

Consider first the regulation of railroads. Over a century ago the railroads were encouraged by both the Federal and state governments, through the grants of charters and franchises, through loans and subsidies, and above all by means of extensive land grants. By the 1870's, however, the practices of the railroads had aroused much public resentment, principally because of their methods of rate discrimination. The railroads could and they often did lower their rates drastically to get new business or to attract traffic away from one another. Large shippers were able to force the railroads to grant them especially low rates, or to pay rebates. As a result, railroad rates were low for some shippers, some classes of commodities, and some cities and areas, while at the same time being high for others. Without much success, the states attempted to regulate railroads in order to meet the demands of farm organizations and of small shippers. Federal regulation began with the Interstate Commerce Act of 1887 which rested upon the commerce clause of the Constitution. The Interstate Commerce Commission (ICC), a bipartisan body engaged in continuous administration of the Act, became the model upon which later public-utility regulation was based. The first task of the ICC was to prevent "unjust and unreasonable" rate discrimination by the railroads although it was not given power to prescribe rates. Discrimination as such has not been eliminated. The rate

per pound for carrying watches is higher than for the transport of gravel. Such discrimination is legally permissible, but it is considered unjust discrimination to charge one shipper of gravel or of watches a higher rate than that paid by others for transport between the same two cities and under the same conditions.[2]

In the period after 1887 much additional legislation was passed by Congress to extend the controls of the ICC and to strengthen its authority; e.g., in 1910 its jurisdiction was broadened to include pipe lines, express companies, and sleeping-car companies. The Transportation Act of 1920 made possible the fixing of the rate level by the ICC, and in addition this Act gave the Commission control over railroad financing and closer control over the services provided by the railroads. Ironically, however, the railroads declined after reaching the peak of their growth at about the time of World War I, so that after 1920 the main task of the ICC was to help the railroads rather than the shippers. The motor carriers (under ICC control since 1935) and air carriers, as well as private automobiles, have provided strong competition for the railroads; except during World War II many railroads have had persistent losses since 1920. Public policy has thus been directed toward increasing their revenues; permission to increase passenger and freight rates was granted in the period just after World War II.

Before the development of the railroad, the Federal and state governments had promoted the building of canals which, except for very special circumstances, have been made obsolete by the railroad. But federal policy has been to continue to subsidize the inland waterways, largely under pork-barrel legislation favoring certain areas. The subsidy policy has been justified on the grounds that waterways have, through their lower rates, afforded a means of holding down railroad rates in the same territories, and that waterway development promotes the conservation of natural resources. The transport companies using the inland waterways are regulated by the ICC.

Ocean and coastal shipping is under the United States Maritime Commission, established in 1936. American merchant shipping had gone into a decline in the middle of the nineteenth century, with the obsolescence of the wooden sailing vessel. In both world wars, large fleets of merchant ships had to be hastily constructed, and after World War I it became a national

[2] In the case of electric power companies, whose rates are also regulated, rate discrimination likewise is important. The rates charged for electrical energy vary with the quantities purchased, with the class of consumer, and with such matters as the time of day (lower rates in off-peak hours) and with the type of equipment. Here too the discrimination must be "reasonable," and justifiable on some plausibly objective grounds.

policy to subsidize a merchant marine. It costs more to construct a vessel in the United States than in foreign countries, and expenses of operation of American vessels are also substantially higher than for most foreign vessels. Two types of subsidies have been granted since 1936: one for the construction of merchant ships, and the other to keep them in operation on the high seas. The subsidies are in principle based on the differences in cost between this and other countries.

Commercial air transportation is regulated under the Civil Aeronautics Act of 1938. The Civil Aeronautics Board controls the economic aspects of the industry, and the Civil Aeronautics Administration administers safety provisions for all flying activities. The air-transport industry has been heavily subsidized, through the construction of airways and of airports by the Federal, state, and local governments, and through payments made by the Federal government (determined by the CAB) for carrying mail. The mail-pay subsidy has been reduced to nominal levels for some of the large air lines with profitable routes; i.e., the payments made to them for carrying the mail are about equal to the costs so incurred. In passing the Civil Aeronautics Act in 1938, Congress clearly intended to foster the growth of air transportation for reasons of national defense as well as to speed the mails. Rate regulation, i.e., of passenger fares and of express and freight rates, has so far been almost entirely nominal. The real problem of regulation has been *control of entry* into the industry. A "certificate of public convenience and necessity" issued by the CAB is required for entrance into the business. The air carriers in existence in 1938 were given "grandfather" certificates; since then there have been hundreds of applications to the CAB for permission to operate, but very few new certificates have been granted.

There is no national transportation policy, no system of integrated regulation for all forms of transport taken together. Each major transport industry is controlled differently. Air and water carriers are subsidized heavily. The extent of the subsidization to motor carriers through the public highway system is a matter of controversy, while the subsidies to the railroads were made in the past. Some rates are regulated closely, some loosely, and some not at all. In general, it can be said that Congress has followed a policy of maintaining a form of competition in transportation. Within each transport industry there are varying degrees of oligopolistic nonprice competition. And there is competition between transports, e.g., between rail and bus, rail and air. This policy of competition is often condemned, on the probably specious ground that integration would achieve important economies of operation.

Those who favor such competition as exists justify it on the ground that each form of transport is under the constant incentive to improve its services.

It must be realized that the transport companies have developed a strong vested interest in the existing controls. This is mainly because of the protection which the control affords against newcomers. The air lines do not object to the regulation of the CAB. On the contrary, they are wholeheartedly in favor of it, because the CAB has protected them from the competition of those who have sought to enter the industry. Efforts to develop a national transport policy fail because of the clash of interests among the transport industries.

In the case of the public utilities other than those in transportation, methods and problems of regulation are similar. Telephone, telegraph, and radio are regulated by the FCC; the telephone industry is also subject to state jurisdiction. The electric power and gas industries are under state public-utility commissions, and are subject to control by the SEC with respect to their corporate structures and finances. Electrical energy transmissions over state borders is controlled by the Federal Power Commission, which also has authority over companies using the energy from water under federal jurisdiction. For these and the other public-utility industries, control means that entry into the industry is limited or impossible because of the public grant of monopoly status, and that service can be extended or abandoned only after permission has been secured. Furthermore, a public utility has a duty to serve all customers, and to serve them equally; it may not refuse service or discriminate among its customers without justifiable reason. Standards of service are also commonly prescribed by law.

Public-Utility Rate Regulation

Rate regulation is the greatest and most difficult of public-utility problems; however, it does not exist in all public-utility industries. In those where some degree of competition exists, rate regulation usually means little more than that tariffs must be filed with a board or commission and that they may not be departed from. But in those industries characterized by monopoly, rate regulation has given rise to great practical difficulties and to a perplexing problem of principle. The practical difficulties are concerned with the exact estimation and calculation of the costs of public utilities, and with the complicated legal proceedings which attend rate regulation. The problem of principle can be stated thus: The value or price of any durable capital good is dependent upon its expected future earning power. The

value of a secondhand truck depends on how much money can be made using a truck. Therefore the value of a capital good is derived from the value of the services produced with it, not the other way around. But public-utility rate regulation reverses economic logic on this matter.

The purpose of rate regulation is to fix a rate or a set of rates which is "fair and reasonable." Left to itself the public-utility company would put the rates too high, from the consumers' standpoint. The public-utility commission is required to act as a kind of substitute for competitive forces, and thus to set rates which compromise the interests of both producer and consumer. Because of the protection afforded the rights of property under the Constitution (the due-process clause), rates must be fair. A fair rate is one which covers costs and which yields a fair profit. A fair profit is one which is comparable, risks considered, to rates of profit earned in competitive industries and which will maintain an adequate flow of new capital to the public-utility industry. But profit rates must be measured against the total investment. Investment means the value of the physical properties ("used and useful") of the public-utility company. The crux of the problem is the value of the properties. To illustrate: Suppose the value of the properties is $10,000,000 and that a fair rate of return is 6 per cent. Then return on investment must be $600,000, and the rates charged the consumer must be high enough to cover other costs and yield this amount of return. If, instead, the value of the properties were $8,000,000, then profits would have to be only $480,000 and the rate to consumers would be lower.

Not too much difficulty has been encountered in determining whether the fair profit shall be 6 per cent, or 7 per cent, or some such figure. Most of the controversy in the courts has centered about the question of according to what principle the *value of the properties of a public utility shall be computed.* From time to time and from case to case the principles advocated and temporarily adopted have been original cost (the actual costs of acquiring and building the properties), prudent investment (original cost as modified by considerations of good business practice), and cost of reproduction (at existing prices). None of these principles is free from objection; each has been the subject of much controversy. There simply is no objectively valid principle of valuation, because it is a question of fixing the value of something, the value of which really depends on rates which are the end product of the regulative process.

Besides these fundamental difficulties, others, such as the reasonableness of operating expenses, add to the complications of public-utility rate regulation. Hence regulation cannot be said to have been successful, although in the face of monopoly poor regulation is better than none at all.

The concept and the institution of the public utility have provided a method of handling the problems connected with those industries which cannot be regulated by the ordinary forces of competition. The monopoly aspect of public utilities has meant not only that some monopolies have been thus controlled; it has meant also that public-utility status has sometimes been sought by industries for the sake of obtaining a legally protected monopoly position. Benevolent regulation is a small price to pay for security from competition.

GOVERNMENT POLICY ON AGRICULTURE

The Federal government has always provided one form or another of aid and subsidy to the farmer. In the nineteenth century the subsidy assured the farmer of free land, and of grants for agricultural education and research. In the twentieth century, special credit institutions were established for farm borrowers, beginning with the Federal Land Banks in 1916. The contemporary farm problem, however, dates from and is mainly a consequence of World War I. To meet the demands of that war, farm production was greatly expanded and in many sections of the country land bought at high prices was heavily mortgaged. Farm prices rose until 1920, and in 1920-21 fell sharply. They did not recover in the 1920's to levels which would have meant an agricultural prosperity commensurate with that enjoyed in most of the branches of the economy at the time. The export markets for American farm products were no longer so favorable as they had been before 1914, and changes in domestic habits of consumption had adverse effects upon the growers of cotton and wheat. In 1929 Congress made an effort at price control. The Federal Farm Board was established for the purpose of stabilizing (i.e., raising) the prices of wheat and cotton. The Board was to have bought these commodities when their prices were low and to sell them when their prices were high. But like many another attempt of the same sort, this one failed, especially because the world-wide decline in demand in the period 1929-33 made price stabilization impossible. The New Deal administration undertook a number of measures to raise farm incomes and to improve the position of the farmer. The farm-debt problem was handled through a temporary moratorium on mortgages, and through federal assistance in refinancing on easier terms and at lower rates of interest. Above all, an effort was made to raise farm prices. The primary purpose of the reduction in the weight of the gold dollar was to bring up the level of farm prices. The Agricultural Adjustment Act of 1933 combined benefit payments with production control; the purpose was to reduce output in order to raise farm

prices to "parity." [3] The growers of farm products other than those covered by the acreage-reduction schemes were assisted by a program of marketing agreements in which the Department of Agriculture sought to negotiate more favorable prices for farmers from the handlers and processors of agricultural commodities. Furthermore, the Federal government bought "surpluses" of various farm products in order to raise or maintain prices and used them as part of the products being distributed among the population on relief. With World War II, farm prices rose steadily. During the war general price control favored farm prices, and, even in the period of high farm prices after the war, the Federal government continued to "support" those farm prices which declined.

In judging the farm policy of the Federal government, the fact that nearly all agricultural commodities are produced and sold under conditions of pure competition must be borne in mind. The individual manufacturer and the retailer have some degree of control, however slight, over their prices. The individual farmer has none at all; and since his is a standardized commodity, the opportunities afforded by nonprice competition are closed to him. Furthermore, farmers' organizations have never had the same opportunities of control which industrial and commercial combinations have had. Thus the farmer has had to ask the Federal government to develop the monopolistic practices he has wanted to imitate. The voting strength of the farm bloc in Congress has been sufficient, during and since the 1930's, to develop a set of controls which has the purpose and to some extent the effect of raising farmers' incomes relative to those of the rest of the population.

THE CONSERVATION PROBLEM

The United States is rich in natural resources, but since many of these are exhaustible there is general agreement that a conservation problem exists.[4] In part, conservation simply means better production practices, as when a farmer plows on the contour to hold erosion in check. In this sense conservation policy entails the spread of knowledge of methods of production in which the interests of the individual and the general interest are in entire harmony. But conservation also means methods of natural-resource exploitation in which the resources are reserved for the future; and here there is often a clash between the general interest and the interests of particular individuals.

It is well known that much of the physical resources of the United States

[3] For a fuller discussion see Chap. 23. [4] See Chap. 3, pp. 68-69.

were squandered recklessly in the nineteenth century. Public policy contributed to the waste. The vast public domain was transferred in large part to private ownership through sale, through grants to the states, to veterans, and to railroads, and through the pre-emption and homestead policies. The policy of alienation in the nineteenth century was in accord with the philosophy of the time which made individual interest of paramount importance.

The conservation movement arose toward the end of the nineteenth century as a result of concern over the consequences of depletion of forests in the eastern United States. Legislation over the years resulted in the policy under which the remaining forest areas owned by the Federal government are treated as permanent resources. The timber in them is viewed as a crop, to be cut only as fast as it grows. The United States Forest Service permits private operators to cut timber on federal land only through methods which are consistent with conservation. A large part of the forest resources, however, is privately owned and is not subject to direct government control.

Conservation of oil resources is another widely recognized problem. The clash between the private and national interests is particularly strong here, because private ownership of land often conflicts with the geology of oil. It is to the immediate interest of each owner of land in an oil field to drill from his own land and to drain as much as possible from the underlying pools. The result has been much avoidable waste of oil and natural gas. Conservation measures have been taken by both Federal and state governments. In the states conservation has been partly successful because it could be linked to price stabilization policies in the 1930's. The major oil companies, fearing the effects on prices of too high a rate of production, were largely instrumental in having state conservation laws passed. These laws provided in the 1930's for reduced rates of production. The result was a tendency toward higher prices than otherwise, injury to the small producers, and some actual conservation of oil. Federal controls have mainly been concerned with the regulation of leases to develop oil sources on land owned by the Federal government.

Soil conservation is a many-sided and difficult problem. The attack on the problem has taken many forms, and education and demonstration, for example, have accomplished much. Flood control and reforestation are related to the conservation of agricultural land, as are the great developmental projects in the Tennessee and Columbia river valleys. The farm programs of the 1930's linked benefit payments to farmers with their cooperation in conservation practices. As in the case of oil, it seems that conservation is particularly successful where it can be tied in with price and production control.

GOVERNMENT OWNERSHIP

Government ownership and operation of commercial enterprises has not developed so far in the United States as in other countries, where as a matter of course governments have long acted as business entrepreneurs. And yet government ownership in the United States is by no means new. The Post Office is a business which could very easily be run as a private enterprise but has always been a government one. Besides the Post Office, the Federal government owns and operates: the Panama Canal, the Panama Railroad Company, and other businesses in the Canal Zone; the Alaska Railroad Company, the Inland Waterways Corporation, the Tennessee Valley Authority and other electric power projects; several lending institutions such as the Reconstruction Finance Corporation and the Federal Land Banks; and housing projects. With the Atomic Energy Commission it embarks on a wholly new field of public enterprise having enormous potentialities for the future. In addition, the Federal government produces hundreds of articles in establishments operated by the Defense and Interior Departments. State governments operate liquor stores, transport systems (e.g., the New York Barge Canal, and the Port of New York Authority, a joint enterprise of the states of New York and New Jersey), and housing projects. Municipalities own and operate electric power companies, water systems, airports, street railway systems, gas companies, housing projects, and similar enterprises.

The causes of government ownership are manifold. National defense is one important factor. It explains the Panama Canal and government ownership of all plants producing fissionable materials. In other countries military considerations have played a much more important role in extending the entrepreneurial functions of government. National emergency brought the RFC and other lending agencies into being during the 1930's. The absence or the failure of private enterprise to carry out some task considered important has been responsible for such public enterprises as farm lending agencies. The facts that electric power production exhibits great economies of size, and that it can be integrated with natural-resource development, have brought government enterprise in this field. Fiscal monopolies—businesses operated by governments for profits—are rare in the United States, being represented by the state liquor-store systems, but they have long been common on the continent of Europe. In some cases government ownership has been adopted because it affords a means of subsidizing particular groups; public housing is an example.

Government ownership as the general or universal method of conducting economic activity means socialism. But government ownership in particular industries in a general context of private enterprise is to be viewed as a method of regulation and as a method of executing public policy.

One of the problems of government ownership is that of method of organization. A business may be operated by a government either directly, as in the Postal system, or through corporations (with varying degrees of autonomy), as in the TVA. The trend in the twentieth century, both here and abroad, has been toward greater use of the quasi-independent public corporation. The advantages of the public corporation are that its finances are separate from those of the parent government and may therefore be viewed more clearly, its management is freer from day-to-day political pressure, its personnel policies can be spared the weight of the usual civil service practices; in brief, it can operate more flexibly than a government department. Between the large management-controlled private corporation and the public corporation, there may not be much difference, except that the public corporation is necessarily less adaptable inasmuch as it operates under a statute outlining its policies and methods, often in minute detail. The public corporation can be too free of political control, i.e., it may carry on activities of which the legislature would disapprove if it were fully informed. There is thus a basic dilemma in government ownership: to be efficient, considerable autonomy of management is usually requisite, yet since government enterprises exist to carry out public policies, they must be subject to as much control as legislative bodies wish to impose.

Whether or not a government enterprise earns a profit is nearly always a matter of minor significance because earning a profit is not usually the purpose of its inception. The real question is whether it satisfactorily performs the functions assigned to it, and whether it is efficient in the sense of good management. Generalization about the quality of management of government enterprises is risky. Some are well managed, some are not. Very often the caliber of management of public enterprises is high when they are new, when enthusiasm to perform an important public task is great, and when capable men can be attracted into public service. But when the public enterprise settles into a routine, when first-class men are drawn elsewhere, quality of management suffers just as it does under comparable circumstances in private business.

Sometimes, government enterprises are in indirect competition with private enterprises. This is true in the electric power and banking fields. Here the questions of costs and of cost calculations emerge as subjects of con-

troversy. Private power companies have complained that TVA's rates are too low because the costs of flood control and of other resource development have not been properly separated from the costs of power production; that not enough of TVA's total costs have been allocated to its output of electrical energy; that TVA is free from some taxes; that TVA borrows its funds at the low interest rates of the United States Treasury. Controversy over questions such as these is inevitable, and is not easily resolved. And on the other hand the lower rates established by TVA actually did cause a reduction of the rates charged by private power companies.

The "yardstick" idea, i.e., that government enterprises ought to be models by which to gauge the social performance of private enterprises, is of little use. Costs between public and private enterprises are not comparable for a variety of reasons, chief among which is that any two enterprises, public or private, are almost always so differently situated in many little respects that cost comparisons do not permit facile judgment as to efficiency.

Intervention in Other Systems

The decisions of producers which determine the composition and the level of national output must always be made in a context of law—in this sense a wholly "free" economy is not to be found in any developed community. Beyond the general rules of social activity established by law there is an infinite variety of conditions possible. On one end, the decisions of private economic agents may be governed wholly by market considerations and each may be free to dispose his income, time, and efforts with no nonmarket influences intervening. This is individualistic, free enterprise. At the other extreme, all decisions directly or indirectly may be made by government. The Soviet Union approaches this limit. The position of the American economy is somewhat removed from the individualistic free-enterprise end of the spectrum, and far removed from that part exemplified by the Soviet Union.

There are many aspects to a contrast of the different national economies and probably the economic contrasts are less important than those afforded by the different political, sociophilosophical, national, and ethnic characteristics. However, the rapid growth of governmental intervention in the United States and the rise of private power groups in business, agriculture, and labor raise the question of contrasts (and comparisons) with socialism, communism, and fascism. Contrasts and comparisons in this large matter may, for present purposes, be limited to the *power to make economic deci-*

sions, recognizing of course that such power rests ultimately on the political character of the economy.

In Chapter 24 it was shown that economic decisions in this country are made predominantly on the basis of private interest. Consumer decisions rest upon individual (and family) preferences,[5] personal income, and the prices of commodities. As the only principal economic agent as yet unorganized, the consumer segment of the economy is mainly representative of the individualistic end of the economic spectrum. For the most part, the consumer is the final arbiter in the economic process, for the bulk of the national output is produced directly or indirectly for consumption by individuals (and families).[6] Producers, it has been shown, increasingly form into large aggregates, either as private associations (e.g., corporations and unions) or as political blocs (e.g., farmers and to some extent wage earners); they remain private but have become less and less individualistic. They serve, however, the individualistic ends of the consumer, although in large part their organizational efforts are intended to gain advantage (1) over competitors, (2) over other parties to the particular productive operation (e.g., employer over wage earner, and vice versa), and (3) over the consumer. Often, the organized efforts of enterprise to overcome competition may result in consumer advantage owing to the economies of large-scale, oligopolistic firms, but this is not an assured result and not necessarily an intended one.

Partly because of producers' combinations and organizations, governmental intervention has grown, although war and other emergencies have also been a major cause. The effect of governmental intervention on economic decisions has been primarily to establish a control over the conditions within which private-producer decisions may be made. The purpose of this control may be viewed as largely to promote *private* interest—to mitigate the conditions confronting one group of economic agents caused by the actions of another, and those general adverse conditions, such as depression, which are beyond the control of any group. In this sense, governmental intervention largely serves private interests, which broadly conceived will include shareholders, farmers, wage earners, business, and consumers.

Despite the growth of organized groups and governmental intervention, economic decisions are in large part *decentralized.* Decentralization is most evident in consumer decisions, which in turn force a large measure of decentralization in the productive process, for the resources essential for

[5] Formed in a context of custom, convention, habit, religion, and law.

[6] The assiduous attempts of producers to sway the consumer do not affect the point made here because the consumer has the choice of being, or not being, swayed.

production must be free to shift into uses required for the satisfaction of consumer preferences. Labor, capital, and enterprise adapt themselves for this purpose in pursuit of their private interests. Thus it may be said the distinguishing feature of the capitalist system is the decentralization of economic decisions, made in response to private purposes. The growth of governmental intervention has been largely in keeping with these private purposes, being limited by the due-process clause of the Constitution.

In contrast, and again in general terms, the other economic systems—fascism, socialism, and communism—are characterized by a subordination of all economic purposes and decisions to state (or governmental) purposes. In both Italian fascism and German National Socialism, private ownership of the means of production was retained substantially as in capitalist countries, but the purposes of the state (military and nationalist objectives) predominated. To the extent that private interests, and therefore private economic decisions, furthered state purposes governmental controls were of no importance; however, when private and public objectives differed, economic decisions were forcibly imposed by public authority. Similarly, in both socialism and communism[7] the power to make economic decisions is vested in government by means of its ownership of the productive facilities of the economy and its political power to make its purposes effective. Thus, in fascism, socialism, and communism, the power to make economic decisions is *centralized* and the objectives of economic activity are politically determined by those who command the governmental apparatus.

The centralization of economic decisions, therefore, enables the economies thus organized to determine (given their resources) the composition of their national outputs and the levels of national output. This is done in accordance with the aims of the state and the energy and enterprise exhibited by the ruling bureaucracy in the management of resources. In the American economy, on the other hand, the composition of national output is determined mainly by the decisions of innumerable producers in anticipation of the demands of consumers, and the level of output is governed by producers' expectations of future demand. Governmental intervention affects both the composition and the level of output but not, except during war, in substantial amounts. Moreover, unlike the other economic systems, governmental intervention involves less the establishment of state aims than the support of private aims by public measures. Thus the contrast between this economy and the other types of economic system is a sharp one; in the others, economic activity mainly serves the state, here, mainly the individual.

[7] Contrasts between these systems will be discussed in Chap. 42.

PART VI

GOVERNMENT FINANCE

CHAPTER 31

Government Expenditures and Budgets

Government finance, or public finance as it is also called, includes not only taxation but also the spending and borrowing activities of government. Moreover, by a long-established convention the terms *public finance* and *government finance* refer to the whole range of economic activities of governments, not solely to their purely financial aspects.

Less than a fifth of the total of goods and services produced in the American economy during peacetime consists of goods and services produced by the Federal, state, and local governments. During World War II nearly one half of the total national output was purchased by the Federal government alone. For decades the *proportion* of total national output consisting of governmentally produced goods and services has been steadily rising. Considering that total resources are divided between the public and private sectors of an economy, it will be seen that the use of material and human resources by government in the production of public services constitutes a diversion of these resources from the private sector. The public sector of the economy includes the Federal government, the state governments, and the thousands of local governments; in the private sector there are business corporations, co-operative organizations, and individual enterprisers. National output consists of the goods and services produced in both sectors of the economy. While they are intermeshed, each sector, however, has its own principles of resource allocation, its own method of determining what

is to be produced, its own method of distributing its output. There are several main points of similarity and of difference between them which require examination.

THE NATURE OF PUBLIC FINANCE

It is commonly said that "private enterprise is carried on for profit, whereas government produces for the needs of the people." The implied contrast is that the former is motivated by selfishness, the latter by altruism—one by individual gain, the other by public benefit. But this contrast is essentially a false one because governmental and private economic activities are here compared at different levels. Both segments of the economy consist of individual units. In private enterprise the units are individuals and business firms; in government there are federal, state, and local departments, boards, agencies, commissions, bureaus, offices, and other units which are not always immediately definable. Thus, in comparing and contrasting public and private economic behavior, units must be compared with units, and wholes with wholes. To say that business works for profit, while government serves the needs of the people, is to contrast units with wholes. Indeed, business corporations attempt to maximize their profits, just as individuals pursue their own advantages, which are chiefly in earning money. But the whole system of business enterprise—which is here called the private sector of the economy—operates to meet the wants of the people.

Private enterprise satisfies wants as it is guided by prices and within the framework given by the distribution of incomes. The entrepreneur can make profits only so long as other individuals voluntarily buy his product; and they buy it only if they want it and have the money to make their wants effectively felt. The public sector of the economy also satisfies the wants of the population, as these wants are determined by representatives in legislatures. The representatives in turn are governed by votes. The units of the public sector correspond in a rough way with the firms of the private sector. Just as the firm pursues its own advantage subject to the condition that it satisfies others' wants, so the governmental unit pursues its advantage subject to the condition that it satisfies wants as determined ultimately by votes. A business product that ceases to be wanted will cease to be produced; a government service that ceases to be supported legislatively will cease to be produced.

Although this process is directly evident in the American society, it does not operate in an immediate cause-and-effect relationship. Some private

producers are maintained by the economy even after they cease to serve wants effectively, and others are prohibited from performing their functions; in government, some agencies are instituted or perpetuated for selfish reasons, others discontinued likewise, against the will of the majority. Thus monopoly in private enterprise distorts the uses of resources in satisfying wants. Political manipulations which impose the will of minorities upon majorities often deflect the use of resources from the majority's wants. Such parallels in the operation of private and public units in an economy cannot be carried too far, of course, since the policies of either vary at any particular time. But it may be seen, at least within the context of a free government and free-enterprise system relatively, that it is essentially untrue to hold that the one sector of the economy is motivated selfishly and the other altruistically.[1]

There are, however, important contrasts between public finance and private finance. One of these has to do with the payment for and the receipt of services. In the private sector, all goods and services are exchanged on an equivalence basis: individuals and corporations buy and sell goods and services, with prices as ratios of exchange and money as the means of payment. All goods and services produced are delivered to consumers who buy the items they want and can pay for. The whole complicated set of processes involved here is individualistic and voluntary. Throughout there is the *quid pro quo:* payment and receipt. Only among friends, members of families, and in truly charitable activities is the *quid pro quo* ignored. Within the government sector of the economy, however, it is different. The receipt and payment for public services by individuals are not on an equivalence basis. Only very exceptionally is there a *quid pro quo.*

Public education may be taken as an example of the receipt and payment for publicly produced services. The services produced in a public-school system are received directly by the pupils, somewhat less directly by their parents, and indirectly by the entire community. (Of course, both the direct and the indirect advantages of education are intended to be produced.) Public education is paid for by taxes. But the children do not pay the taxes, nor do their parents *because they are parents,* nor do employers who benefit from the training afforded in public schools. Taxes are not paid in a direct and particular exchange for government services enjoyed by the individual taxpayers. If an individual pays a private person or firm a hundred dollars, he presumably receives or will receive a hundred dollars' worth of commodities or services; moreover, he pays voluntarily. If the same individual

[1] See Chap. 30, pp. 664-66, for a discussion of public agencies in competition with private agencies, as an illustration of similarity between them.

pays a hundred dollars in taxes, his claim upon the services of government is not thereby affected; his payment of taxes is compulsory. Once a tax has been decided by a legislature, the persons concerned have no other choice than to pay.

The expenditure of government funds and the receipt of the funds by taxation are thus two quite separate processes. There is no *quid pro quo* for the individual. Still other factors peculiar to the consumption of publicly produced services may be noted. Some public services are compulsorily consumed. Elementary and secondary education, like public health measures, are required to be consumed; the services of a state university, on the other hand, are voluntarily enjoyed. Furthermore, many public services are consumed collectively. The provision for law and order and for national defense necessarily affects every individual. It is impossible to determine, in fact, just what amount of public services are consumed by any one individual. In contrast, the services (and goods) provided by private enterprises are always sold on a unit basis and bought by individuals. But how can a unit of national defense or of law and order be measured? Such services cannot be broken up into pieces or items, like private services, and calculated on an individual basis.

The decisions of private enterprises turn upon a clear-cut and simple standard of performance and success. Whatever the criticisms of profits and the profit motive, it would be difficult to deny the simplicity and the effectiveness of the quantitative standard of profits in a free enterprise economy. Profits mean success: the higher the profits, the greater the success. Private enterprises can be quickly and easily compared with one another by reference to their profits and losses. In the public sector of the economy, however, there is no single standard of performance and efficiency but instead a multiplicity of standards. Is the national defense adequate? Does the school program fulfill its purposes? Should public works be increased? These and hundreds of similar questions cannot be answered finally and to everyone's satisfaction. But they do in some way have to be answered, and as a result the debate over government expenditures and tax policy goes on endlessly.

The expenditures which any individual can make depend upon his income. Hence, income can be thought of as coming first, and expenditures second, in the individual's decisions. With governments, however, it is the other way around. Expenditures are decided upon first, then provision is made to raise the necessary revenue. This can be done because governments have the power to raise funds by compulsion, i.e., by taxation. The principle that raising revenue follows the determination of expenditures

requires, however, important qualification. The ability of local governments, such as cities and counties, to increase tax rates is in practice usually very slight. Hence local governments are generally circumscribed as to the services they can provide. The revenue-raising powers and therefore the range of expenditures of the state governments, while relatively greater than those of the local units, are also limited. The revenue-raising *powers* of the Federal government are for most practical purposes unlimited, as is proved in time of war; hence, the Federal government has relatively the widest latitude in setting expenditures. But while taxes can always be increased, and additional funds can always be borrowed, Congress usually has some notion as to how much taxes the population will tolerate. The decision, therefore, that taxes ought not to go above some level, is always at any given time a factor limiting expenditures except in times of national emergency.

Still a further difference between public finance and private finance, which at first sight seems to be superficial but which is actually major, is from their different methods of accounting. The various governments (in the United States, at least) do not compile and publish balance sheets, i.e., statements of assets and liabilities. Conceivably, the balance sheet of a government should show debts and property owned. For the Federal government, the major debts would be two: currency (including gold certificates) outstanding, and interest-bearing securities outstanding. The latter is usually called the national debt. The assets, or property, of the Federal government consist of such things as gold, silver, land, buildings, and equipment of all kinds. So far, no one has ever attempted, for the purpose of a balance sheet, to arrive at the value of all of the property of the Federal government. It is important to note that if such a figure were found and used in government accounting, it would probably alter the significance of the national debt to most people. In private accounting, debt is usually reckoned as *net* debt, i.e., gross debt minus assets. But governmental debts are always given simply as *gross* debts. This fact is commonly overlooked, with the result that the national debt is viewed with a somewhat distorted perspective.

It is obvious that various interpretations of these relations and contrasts between private and public sectors of the economy make public finance a field of economics in which there are perennial controversies—some without prospect of solution. If a governmental balance sheet were established in accordance with commercial habits of thought, it might narrow somewhat the range of the controversies over the single issue of the national

debt. A similar issue is found in the way governments calculate their expenditures. If an office building is erected and is intended to last twenty years, private business will—in principle—spread the cost of the building over the twenty years. Its expenses for each year will include one twentieth of the total cost of the building. But the practice of the Federal government is to report all expenditures as they occur, regardless of the purpose of the expenditure. If it takes a total of two years to erect the office building, the Federal government's expenditures for those two years will show the total cost of the structure. The effect of this practice is to make government expenditures look larger in some years and smaller in others, as compared with the bookkeeping habits of commercial enterprises. Here, too, a bookkeeping difference which is not generally understood adds to the flames of controversy.

THE GROWTH OF PUBLIC EXPENDITURES

Public expenditures have grown steadily in volume since the founding of the nation. During wartime there have always been large increases, and only in the periods after major wars have there actually been decreases. Even in these periods the declines in expenditures were to levels much higher than had existed in the corresponding prewar periods. The expenditures of the Federal government alone for a few selected years are shown in Table 62. The years (fiscal years ending June 30) have been selected so as to indicate the effects of major wars, the declines after wars, and the growth in peacetime.

When state and local expenditures are added, the totals of public expenditure become much larger, especially for the earlier years. In the period before World War I, in fact, federal expenditures were only between a quarter and a third of total governmental outlays. More than half of public expenditures were made by local governmental units. However, in the 1920's and 1930's, federal expenditures were a higher proportion of the total. Just before World War II, federal spending was about one half of the total; the other half was accounted for by state and local expenditures, those of the states being less than those of the local units. Table 63 gives figures for state and local expenditures for several years since 1920.

Although it may be said that federal expenditures have grown over a long period of time because of the occurrence of major wars, state and local expenditures, which are much less affected by war, have nevertheless persistently increased.

Table 62 EXPENDITURES OF THE FEDERAL GOVERNMENT FOR SELECTED YEARS

(in millions of dollars)

Fiscal year	*Expenditures*
1791	4
1830	15
1865	1,298
1867	357
1903	475
1916	724
1919	18,952
1920	6,403
1925	2,464
1930	3,237
1935	7,010
1940	9,445
1945	100,405
1947	42,505
1948	37,405*

Source: In part, the figures are from those presented in Harold M. Groves, *Financing Government*, rev. ed., Holt, 1945, p. 469.

* As estimated in January, 1947.

Table 63 EXPENDITURES OF STATE AND LOCAL GOVERNMENTS

(in billions of dollars)

Calendar year	*State expenditures*	*Local expenditures*
1920	0.9	3.3
1925	1.5	5.6
1930	2.2	7.2
1935	2.2	5.5
1940	3.6	5.6
1943	4.3	4.6

Source: Adapted from *Public Finance and Full Employment*, Postwar Economic Studies No. 3, December, 1945, Board of Governors of the Federal Reserve System, Washington, D. C., 1945, p. 115.

THE SCOPE OF GOVERNMENT EXPENDITURES

For the Federal government the consequences of past wars and the preparation for future wars have always been the major activity. The national defense expenditures are those of the Department of National Defense, and may be taken to constitute the preparations for (or prevention of) future wars. Aids to veterans and interest on the national debt are the expenditures caused by past wars. Although not all of the national debt in the past was contracted in war periods, since, for example, there was a large increase in the federal debt in the 1930's, it may be said without serious qualification that the interest payments on the national debt are for practical purposes a war expenditure. Hence, the three main classes of federal expenditure are (1) national defense, (2) aids to veterans, and (3) interest on the national debt. These have constituted about four fifths of all federal outlays. The only exception to this was the decade of the 1930's, when the ratio of expenditures in these three classes to the total federal outlays, fell to between a half and a third of the total. The explanation is that the depression of the 1930's caused unprecedented emergency expenditures by the Federal government for public works, relief, and aids to farmers.

National defense expenditures have in the modern period shown a general tendency to increase in peaceful years because of the mechanical—now physical and chemical—complexity of recent warfare and because of the increasingly important position of the United States in world affairs. From 1920 to 1930, War and Navy Department expenditures were about three times as large as from 1900 to 1910; expenditures in the earlier period were about five times as great as in the decade from 1870 to 1880. The budget for national defense for 1947-48 was 10.4 billion dollars. This figure was nearly three times greater than that for the comparable year following World War I, 1919-20.

The only major federal expenditure apparently capable of being decreased is the interest on the national debt. It has been the tradition of the United States to repay its wartime debt. As repayment takes place, the total of interest payments declines. This source of decrease, however, has always been more than offset by the growth of expenditures for veterans aid after each war. The tendency has been for legislation to provide for the veterans of each war progressively more liberally. In each of the decades after both the War between the States and World War I, total payments increased, instead of diminishing. The budget for veteran aid for

1947-48 was 7.5 billion dollars, which was about 22 times greater than the amount provided in 1919-20.

The civil expenditures of the Federal government, although a small proportion of the total, have also grown steadily in the course of national development. New functions have been steadily added, and the performance of old functions has been extended. In the 1930's, the RFC, the AAA, the various relief and public works agencies, the Social Security Administration, and other agencies created in that period, accounted for a sixfold expansion of the nonwar federal expenditures.

State and local governments have no one major item of expenditure, such as war entails for the Federal government. Table 64 shows the expenditures, by function, of the state and local governments in 1941. About three quarters of all the expenditures of the state governments is accounted for by

Table 64 **EXPENDITURES (EXCLUDING DEBT RETIREMENT) OF STATE AND LOCAL GOVERNMENTS IN 1941, BY FUNCTION**

Function	*Per cent of total*
Schools	25.9
Highways	17.2
Old-age and unemployment ins.	5.0
Welfare	12.6
Health and hospitals	6.2
Corrections	1.6
Sanitation	3.0
Police and fire	7.3
General administrative, legislative, and judicial	8.2
Interest	5.8
Recreation	1.9
All other	5.3

Source: Public Finance and Full Employment, op. cit., p. 104.

highways, social welfare and corrections, and education. The proportion of state funds spent on highways rose sharply during the 1920's, the absolute amount expanding by the hundreds of millions. Moreover, the 1930's imposed a heavy burden of relief expenditures on the states. State expenditures on education are primarily in the form of grants to local governments. These grants are intended mainly to raise the minimum standards of education, and in recent times the tendency has been for a pronounced increase. In addition to highways, welfare, and education, the states spend money

for conservation, for health and sanitation, for recreation, and for police systems.

The local governments—counties, cities, townships, school districts, and other local units—also spend most of their revenues for education, welfare, and streets and highways. Besides these items, other important ones are "public safety," health, sanitation, and recreation.

The expenditures of the Federal, state, and local governments are directly interrelated, for the Federal government makes grants to the states, and states make grants to local governments. Federal grants to the states are for such purposes as highway construction, vocational rehabilitation, old-age assistance, aid to the blind and to dependent children. State grants to the local governments are largely for education, highways, and welfare.

CAUSES OF THE GROWTH OF PUBLIC EXPENDITURES

As has been indicated, the greatest single cause for the growth of public expenditures has been war. The War between the States and World Wars I and II caused federal expenditures to increase during wartime and postwar periods above the combined expenditures of the state and local governments.

Among the other factors making for the expansion in public spending is growth of population. As the country has grown, so have the expenditures of governments. Per capita expenditures, however, have grown too. Technology has been a significant force in this increase. Even if the United States had not engaged in war, a steadily developing technology would still augment the cost of national defense preparation. Decade by decade, each member of the armed forces has had to be outfitted with more elaborate and more expensive equipment. Nonwar expenditures are likewise affected. A mile of highway costs more than it did a generation ago. Both labor and material costs have gone up, but highway-building technique has improved. The narrow two-lane road was cheaper but also poorer than the four-lane divided speedway with clover-leaf intersections. Life in the large city has become more complex. Similarly, the maintenance of the health and well-being of the city inhabitant requires a larger expenditure per capita than in a small town. Urbanization as a social phenomenon has always tended to raise the levels of local public expenditures. Water, sanitation, health and welfare, transportation, recreation, police and fire protection are all highly organized and expensive activities of the modern city. In the village of the nineteenth century the citizen took care of his own water supply and sewage disposal; he looked for no public health assist-

ance or transport facilities; he had not heard of organized recreation and expected but little police protection.

The idea of a "rising standard of living" is familiar enough. The notion of a rising standard of living may be applied to public finance. Total consumption consists of both privately and publicly produced goods and services. Just as the standard of living for privately produced goods has risen over the decades, so has the standard of living for public services. A public-school system today is expected to provide better facilities and instruction than a generation ago; children are expected to receive more years of schooling. Most local governments are expected to provide public parks and other recreational facilities. The motorist of today wants not only a better automobile but also a better highway network than were considered satisfactory in 1920. The informational services of government have more demands upon them than formerly. Examples such as these could be added indefinitely.

It is readily discernible that the growth of the activities of government and the attendant increase in public expenditures have been a central factor in the economic development of the nineteenth and twentieth centuries. Students of public finance have enunciated a "law of increasing public expenditures," according to which the public expenditures of a nation grow—both intensively and extensively—with the industrialization of a nation. It is quite true that public expenditures have in the time of advancing technology grown intensively and extensively, i.e., more has been spent on old functions and new functions have been added. But industrialization is only one factor and it is probably a superficial one at that. The development of industrialism is historically complemented by the strengthening of humanitarian impulses, the extension of democracy through political suffrage, and the broadening of concepts of government; all have been part of social development of the past century and a half. Step by step more tasks have been assumed by government, step by step new problems have emerged to be undertaken by processes of government. In modern times, then, ideas on government and social behavior have tended toward increased government influence—whether for good or bad—on national life. Industrialism has created new problems and made old problems more urgent within the scope of this trend.

The depression of the 1930's presented to the United States its greatest test of government responsibility in the conditions resulting from large-scale unemployment. The Federal as well as the state and local governments were required to establish agencies and measures to cope with the emergency. Subsequently, the functions of government undertaken during

this time were retained and some became almost permanent additions to national policy. The RFC, for example, was established in 1932 to meet the emergency presented by railroads and banks on the verge of collapse. The emergency of 1932 passed, but others took its place, so that in a sense the RFC has thrived on a series of financial emergencies since 1932. The emergency legislation of 1933 to help farmers by raising farm prices was followed by a development of controls over agriculture which give every indication of continuing, in some form or other, for an indefinite period. It must be allowed of course that the portent of war followed closely on the trial of depression and hence continued government action was necessary, but other periods of increased government activity, such as the periods from 1860-65 and 1916-20, resulted in the retention of governmental authority in many fields. Once a policy is established for an emergency it tends to create a changed condition in the relation between private and public pursuits which favors its retention.

THE ECONOMIC EFFECTS OF GOVERNMENT EXPENDITURES

The effects of government expenditures upon the private enterprise sector of the economy vary. To appraise these effects, a distinction between *transfer expenditures* and *real expenditures* of the government must be made. Transfer expenditures are those which accomplish a redistribution, i.e., a transfer of private income, whereas real expenditures are those which involve the direct use of resources by public agencies.

Consider first the transfer expenditures. Examples are interest on debt, payments to veterans, social security and relief payments, and benefit payments to farmers. Suppose that the funds for these expenditures are all raised by taxation. What takes place, then, is a transfer of income from taxpayers to the owners of government securities, to veterans, to the beneficiaries of social security and relief policies, to farmers, and to other groups in the economy. What are the consequences? In the simplest interpretation, the effect is that the taxpayers consume less, the groups benefited consume more, and total consumption is unchanged. The cost of administering such transfers of income are negligible in comparison with the amounts involved. If total consumption is unchanged, it follows that the total private production of goods and services is unaltered. The loss of demand among taxpayers for goods and services is balanced by the demand created by veterans, the aged, and the other groups. It is recognized that in all probability the allocation of private production is different because of the existence of transfer expenditures. The commodities the taxpayers would have

bought are likely to be different ones from those purchased by the holders of government bonds, veterans, etc. But beyond this, transfer expenditures have no effect upon the private sector of the economy.

Suppose, however, that 10 billion dollars of taxes reduce taxpayers' consumption by only 7 billion dollars, the other 3 billions being a reduction of the current savings the taxpayers would otherwise have undertaken. Suppose further that the 10 billion dollars is paid out to persons who spend the entire 10 billion dollars for current consumption. It follows that total consumption in the economy over a given period will be larger, assuming that the other factors influencing consumption do not offset these effects. Conversely, total consumption could conceivably be smaller—if the recipients of the expenditures spend less for current consumption than the taxpayers would have otherwise spent. The effect upon consumption, therefore, depends upon who the taxpayers are and who the recipients are and upon their respective marginal propensities to consume.[2]

A further effect of transfer expenditures relates to investments. Suppose that part of the taxes collected are funds that would otherwise have been invested, and that part of the transfer expenditures (say to large holders of government bonds) go to persons who subsequently make investments. In this case the direction and volume of new investment will probably be different from what it would have been had there been no transfer expenditures.

Expenditures of the transfer type are not necessarily derived from funds obtained by tax collections. The funds may be borrowed, in which case the transfer is from the lenders to the recipients of the expenditure. In this instance the same effects on consumption, as have been described for tax-funds expenditures, may occur. If the funds are borrowed from commercial banks, there results an addition to the money supply.[3] Transfer expenditures in this case would represent no diversion of money income, but rather an addition to the income stream and an expansion of output; real resources could be diverted as the additional money income is spent.

In contrast to the transfer expenditure are real expenditures, which result (under conditions of full employment) in a direct diversion of resources from the private to the public sector of the economy. In making a real expenditure the government does not transfer funds from one group to another, but raises funds (by taxation or borrowing) and carries on activity itself by means of employing labor and buying materials. Public works are a prime example. In a period of full employment, the public works projects of Federal, state, and local governments result in a diversion of construction

[2] See Chap. 6, pp. 142-44.

[3] See Chap. 10, pp. 242-46.

labor and of building supplies and equipment—to the government and away from private employment. Assuming that the expenditures are met from taxes, the effect of real expenditures is, therefore, to reduce consumption of privately produced goods and services and to increase the consumption of publicly produced services. The diversion effected by real expenditures is clearly of a different sort from that caused by transfer expenditures. The latter bring about a rearrangement of private consumption, while the former substitute public for private production. In a period of heavy unemployment, when real expenditures are made from borrowed funds which would otherwise be idle, the real expenditures will increase total (private and public) output. This assumes that the real expenditures by themselves do not indirectly depress private output. In a period of full employment, however, government and private individuals and enterprise compete for resources, in a struggle in which the government has a strong advantage in its power to compel by means of taxation.

To recapitulate the effects of public expenditures on private economic activity:

1. Transfer expenditures may increase or decrease private production and consumption, depending on from whom and to whom income is transferred.
2. Different goods are produced for individuals from those which would be produced in the absence of transfer expenditures.
3. Real expenditures cause a decrease of private production and consumption, except under certain conditions prevailing in depression periods.
4. Private production is affected by real expenditures through their indirect effects on certain industries. (The costs of private housing construction probably would be higher because of a large public works program.)
5. The effects of real expenditures on prices and wages may spread far beyond the industries and areas directly affected by government expenditures. Hence the reallocation of private economic activity caused by the existence of government is probably much greater than is commonly realized.

The Productivity of Public Expenditures

Government expenditures are so frequently referred to—without qualification—as "unproductive" that it is necessary to examine the question of

their productivity. Demands for tax reduction made in the press and in Congress are usually accompanied by charges or insinuations on the allegedly inherent wastefulness of government spending. But what does "productive" mean? In a business context, a productive expenditure by a business enterprise is one which results (or is expected to result) in revenue larger in volume than the expenditure, i.e., a profitable expenditure. The definition of productive, in a general economic sense, is utility-creating. These two definitions are not identical. The former compares expenses with receipts, while the latter looks only to the effects of expenditures. There are some public expenditures which do directly bring in revenue and which are productive in the business sense. A toll bridge and a municipal swimming pool with admission fees are examples, as is the Post Office Department. Such activities and projects may even bring in revenues larger than the outlays and are thus "productive." Other types of expenditure may somewhat indirectly increase tax revenues: an improved highway system could increase gasoline tax revenues. In this way, then, such expenditures are also productive. So too, but less directly, are expenditures for vocational education and for conservation: not only is the efficiency of resource utilization improved but also at some future time tax revenues will probably be larger. The real point of productivity—in the sense of bringing in revenue—is that it is not the essential criterion by which to judge the desirability of public expenditures. This is easily illustrated by reference to national defense and to the maintenance of law and order. Only anarchists would deny the desirability of expenditures for such purposes. But such expenditures are not productive in any commercial sense, unless the customary meanings of words are to be grossly distorted.

If, now, the term *productive* is given its more usual meaning in economics —that which is utility-creating—it is clear that public expenditures which create utilities are productive. The utilities created by governments are of the most diverse types. Those who deny the productivity of particular government expenditures are in effect denying that the activities in question possess utility and their charges should be examined in this light.

Waste is another matter. Waste is the application of resources to unimportant uses while more important uses remain unfilled. It would be wasteful to spend a million dollars in improving a secondary road when improvement of a main highway is more important. Similarly, it would be wasteful to construct a school auditorium if classrooms are needed more urgently. But the difficulty is the measure of agreement on which are the more important public purposes and which are the less important. Here again is a source of controversy over public expenditures. Where there is

general agreement on what is more important, it is easy to identify possible waste of public funds. But where there is a majority and a minority opinion, the one group will find a given expenditure necessary in the public interest, while the other group will denounce it as wasteful.

The Determination of Public Expenditures

What is the essential criterion in the determination of public expenditures? This question must be answered by examining two factors in public finance: (1) *the determination of the total volume of public expenditures;* (2) *the allocation of public funds to various possible directions of expenditure.*

Some economists have attempted to establish a line between the public and private sectors of the economy by reference to comparative marginal utilities. The total population is imagined as deciding whether or not a billion dollars would increase utilities enjoyed more if spent on public services rather than on privately produced services. Presumably, there is a preference for the utility to be gained from one type of service rather than from another. Thus the division of all economic activity into two segments, public and private, reflects the decisions of the people as to comparative marginal utilities of public versus private expenditures. However, serious objections can be posed to this explanation of how the total volume of public expenditures is determined. Most important of all, it overlooks the existence and opinion of minorities. Legislative government is designed to express the will of the majority, but usually there is a dissenting minority, often a large one. Moreover, if representation is not in proportion to population, the minority may have the power as representatives of the majority to spend (or not to spend) government funds, although the entire population is required to abide by the decision, to pay taxes, and to countenance the volume of expenditure.

Earlier it was said that the payment for and the receipt of government services are not made on an individual exchange basis. Because they are not, the comparison of the marginal utilities of public services with the private services otherwise purchasable with tax funds is not a very helpful one. It is not wrong, but its applicability is quite limited.

It is better to explain the volume of government activity, in comparison with that of private economic activity, by examining its cumulative growth over a long period of time. Although public expenditures per capita have grown faster than national output per capita, and although there has therefore been a substitution of public for private services, the growth in any one year has been relatively small. War years, and other years of emergency

such as 1933, are of course important exceptions. The great bulk of expenditures in any one year is continued from the previous year, and also the taxes which support them. Most of the government activities performed in any one year are carried over from previous years. Private economic activity is therefore adjusted, in any one year, to a given volume of public economic activity, and is therefore also adjusted to the effects of both transfer and real public expenditures. It follows that the volume of public expenditures about which a new decision has to be made in any one year is relatively small. The question for legislatures to decide is fundamentally whether the benefits of the proposed new expenditures exceed or do not exceed the benefits from leaving the funds in the hands of individual taxpayers. The bulk of expenditures are governed by accepted past legislation. In this limited way there is a balancing of utilities at the margin, but the balancing is done by elected politicians through the complicated processes of democracy.

The factors which govern the *allocation of the total of expenditures* among the various objects of expenditure are highly diverse, and do not correspond to the relation of factors which allocate resources in private economic activity. To begin with, expenditures are determined by political decisions, which are presumably majority decisions in which some compromise of issues has been necessary. The complex processes of political action can be simplified for the purposes here by saying that, at the highest level, total expenditures are divided among half a dozen or so major purposes. This division or allocation reflects the dominant political philosophy of the administration as well as the aims of the strongest politicians. On the next level, each of the major groups of expenditures must be further divided into its component parts. Thus, for example, once a total appropriation for highways is arrived at, this total must itself be divided into smaller amounts for the various stretches of highway being considered. At this level expert opinion (of highway engineers) is more important, and political decision is less important. The division of a highway fund could be carried out by reference to rational and objective criteria, such as statistical counts of traffic flows, types of traffic, population to be served, exact cost calculations, etc. A public health appropriation can also be split into many small parts, on the basis of purely medical considerations. Each government activity has its own experts, who may be very influential in making final detailed allocations.

Now, these considerations point to one highly significant aspect of the nature of government expenditures. For each distinct type of governmental activity there is a different set of criteria according to which an expenditure

fund is allocated. Because the benefits of government expenditures are generally not calculable in money, they must be assessed in other terms. Defense, conservation, health, education, housing, law and order, and the others must each be appraised not only by technicians in these various fields but by the public in general. The endless controversy over objectives and methods characterizing public finance arises because there can be little agreement on standards. Is the maintenance of public health centers more important than military expansion? or highways? or atomic research? Each allocation of resources has its own value and could perhaps be balanced with others by an individual in possession of adequate information, but hardly by an entire group or population who are free to express differences.

The role of the expert in allocating a given class of expenditures among detailed purposes has been mentioned. The position of the expert—on education, conservation, flood control—is of course not limited to this function. The expert may make his recommendations at the highest (political) level to a legislative committee. Commonly, too, the expert seeks to gain support from the general public to induce the authorization of additional funds for the expert's type of government activity. The expert nearly always wants more. By definition, an expert is a specialist in a field; necessarily, he regards it in a distorted perspective. The politician, on the other hand, must consider many things; he must be sensitive to pressure from many different sides, and must weigh the conflicting claims of many different types of experts. The politician's function has been said to be that of telling the expert what the public will stand; but in any event, the politician stands between the voter and the expert public official. Often, too, the politicians busy themselves with details of expenditures. Roads were mentioned as an example of where objective criteria can be used. But many a road has been built where the politicians rather than the highway engineers wanted it. The allocation of public expenditures among various purposes, then, is carried out through an admixture of expert and of political decision. The total volume of public expenditure is the result of the decisions of Congress, of the state legislatures, and of the legislative organs of the local governments. Each one of these legislative bodies considers the merits of new (or decreased) expenditures against the advisability of additional (or smaller) taxation.

The volume of government expenditure, especially that of the Federal government, affects the level of national output. This was explained in Chapter 6; the policy aspects of this problem will be further discussed in Chapter 33.

PUBLIC REVENUES

The major revenue of government is taxation. Even when, as in wartime, a large part of total revenue is obtained by borrowing, the borrowed funds may be treated as deferred taxation. Interest is paid on government securities, and it is always intended that interest payments will be made from tax revenues. The payment of principal (at least in the United States) is also invariably intended to be made from future tax income. But besides taxes and borrowing, both of which are to be discussed in the next two chapters, there are several other sources of government revenue which should be considered at this point.

The Federal, state, and local governments derive small amounts of revenue from fines and penalties, from gifts, from rents and royalties on public lands, from interest on government loans, and from fees and licenses. The sale of surplus war property brought in some revenue to the Federal government after both world wars. Before the War between the States the sale of federally owned land to settlers and speculators was made for the purpose of obtaining revenue, as well as to encourage settlement of new areas.

The commercial or quasi-commercial enterprises of government also yield some receipts, and in so doing raise important questions of policy, some of which have been discussed previously.[4] The Federal Post Office Department and TVA, state liquor stores, and municipal power companies are examples. The revenues from these agencies are similar to those of private business enterprises for they also are derived from prices charged to individual consumers. A major problem of policy is whether to set the prices of public enterprises at levels which will return a profit, or will equal cost, or will yield a loss. If a municipal power company is operated at a profit, the consumers of electrical energy contribute to the cost of local government, and for most purposes the profit can be considered as a tax levied upon them. There is no doubt, for example, that the profits of a state liquor store system are intended to be a tax on the consumers of alcohol, who represent special consumers of a commodity deemed to be unessential and possibly harmful. But there may be some doubt in the case of power company profits, since one must justify the singling-out of consumers of electrical energy (as consumers and in proportion to their consumption) as a group to be taxed. In other words, without a strong and independent reason of public policy, there is no basis for operating a public enterprise at a profit simply for the sake of the profit. The same observation applies to public enterprises oper-

[4] See Chap. 30, pp. 664-66.

ated at a loss. Unless there is some reason of public policy to subsidize the consumers who benefit, it is difficult to justify the losses. In most cases, the prices received by public enterprises are fairly close to unit costs, so that little or no net revenue is realized by the operating government. Post Office rates illustrate another facet of public prices and public policy. Postal rates are set by Congress, mostly so that total revenue will be fairly close to total costs. But the rates for some of the postal services are patently below the costs of the services. An example is the rate for newspapers and periodicals —a low rate justified on the ground that the arts and literature are thereby promoted. This may be so, but no doubt the influential position of the press has also had something to do with this decision of Congress. The air-mail rate has also been set below the costs attributable to air-mail service, with the avowed purpose of promoting air transportation.

The Federal government enjoys a steady but quite modest revenue from "seigniorage," the profit from making coins. Dollar devaluation in 1934 brought in a substantial "increment." [5] Similarly, the addition of new currency by the Treasury is a source of revenue. It must be added, however, that these latter types of federal revenues have been more than offset since 1934 because of the expenditures on silver which the Treasury has had to make under the Silver Purchase Act of 1934.

Special assessments are a form of revenue to local governments. Special assessments are imposed on property owners where property is improved by the construction of streets and sidewalks and other works. The property owners are usually able to vote for or against the improvement and its accompanying special assessments; if a majority approves, the special assessment falls upon all property owners. In principle, the amount of the special assessment paid by each property owner is equal to the increase in the value of his property attributable to the improvements. Thus, a special assessment differs from a tax. The tax is a compulsory payment for the general welfare not recognizing in particular the welfare of the individual taxpayer, while the special assessment is a compulsory payment by certain individuals for an improvement which indirectly contributes to the general welfare but which directly benefits the assessment payer.

Public Budgets

A government budget is a method of organizing expenditures and revenues. More particularly, the budgetary process involves the planning, comparison, and co-ordination of the several classes of expenditures and

[5] See Chap. 11, pp. 256-59.

revenue, both in the executive and in the legislative branches of government. In the United States, the preparation of public budgets is a relatively recent development. In the nineteenth century and in the early part of the twentieth century, financial administration in the Federal government and in the states was haphazard; there was no rational procedure for considering and managing the finances of the governmental units as a whole. The present budgetary system of the Federal government was established by the Budget and Accounting Act of 1921. This legislation provided for two financial committees in each house of Congress, instead of the large number which had previously existed, and for the General Accounting Office under a Comptroller General, and the Bureau of the Budget. Until 1939 the Bureau of the Budget was part of the Treasury Department; since then, however, it has been a part of the executive office of the President.

There are now two major types of government budgets: (1) the executive budget and (2) the joint budget. The executive budget is a financial plan drawn up by the executive branch of government, which is presented to the legislature in such a way that the budget is either accepted as a whole (with perhaps only minor modifications) or rejected. This system is used in British countries where the executive (the prime minister and his cabinet) is part of the legislature. The great advantage of the executive budget is that an integrated budget is possible, from the initial planning right through to the actual making of the expenditures and the collection of the taxes. Some of the states in this country have modified forms of the executive budget, but the principle of separation of powers inherent in United States constitutional procedure serves to prevent the governor and the state legislature from direct joint action.

The budget of the United States is a joint budget. The budget is prepared by the President and is then presented to Congress, which makes the actual expenditure and revenue decisions,[6] so that there is a joint responsibility. The Bureau of the Budget collects from all of the agencies of the Federal government their estimates of expenditure requirements for the coming fiscal year (beginning July 1). It is to be expected that the head of a government agency will view the functions of his own agency as all-important and will therefore ask for a relatively large amount. The Bureau of the Budget, however, reviews all requests critically and may decrease or increase single estimates. The Bureau then draws up a tentative plan of expenditures which it submits to the President, who meanwhile obtains the advice of the Secretary of the Treasury on revenue, i.e., on estimates of tax revenue and on borrowing requirements or debt-retirement plans.

[6] Subject of course to presidential veto of bills which incorporate them.

With the help of his advisers the President then prepares a "budget message" which is made public and sent to Congress each January. The budget message is therefore an integrated plan of the executive branch. In the months after the receipt of the budget message in January, Congress debates over the expenditure appropriations and tax measures. Under the Legislative Reorganization Act of 1946, provision was made for a joint committee of the Senate and the House to meet with the two financial committees of each of them and to prepare a "legislative budget" as a recommendation to Congress. The final result of the actions of Congress may bear little resemblance to the budget message of the President.

The federal budget process as a whole does not represent the execution of a carefully conceived plan. Many of the estimates which go into the President's budget message to Congress must be prepared two years in advance of actual expenditure or receipt of particular funds. Thus even if Congress adopts the President's recommendations, large-scale mistakes can be made, and it may be necessary to pass "deficiency appropriations" if expenditures have been underestimated.

The Comptroller General is an official appointed by the President, but he is responsible only to Congress. It is his duty to see that the funds voted by Congress and spent by the various federal agencies are actually spent in accordance with the will of Congress. The Comptroller General presides over the General Accounting Office, which also handles claims against the Federal government and audits the books of federal agencies.

SUMMARY

The economy is divided into two sectors: private and public. Public finance differs in some important respects from private finance, notably in the power of public bodies to raise revenue by compulsion, and in the lack of equivalence for any one individual between his tax payments and the public services he enjoys. Lacking the profit principle, governments must apply other standards to judge their activities, and lack of agreement on these is the reason for the endless controversy surrounding public expenditures and taxes. Public expenditures have been growing, both absolutely and relatively, both extensively and intensively. For the Federal government, war has been the principal cause of expenditure, and for the states and local governments, highways, education, and welfare activities. Public expenditures can be divided into two types. One is the transfer expenditure which effects a redistribution of income, and thus a reallocation and possibly increase or decrease of private consumption. The other is the real expendi-

ture which results in the use of resources by government and their withdrawal from private utilization. The productivity (in the commercial sense) of public expenditures is not the most important criterion by which to judge them. Rather, explanation for the allocation of total resources between public and private sectors, and of public expenditures for particular purposes, is found in the factors which guide legislatures. Public revenues consist of taxes, proceeds of borrowing, and a series of minor items. Budgets co-ordinate revenues and expenditures, but because of separation of powers between branches in the American form of government, the budgetary process is rarely exact.

CHAPTER 32

The Tax System of the United States

The tax system of the United States consists of the complex of taxes imposed by the Federal, state, and local governments. The system is here treated by describing the establishment and effects of various taxes. In the following chapter an analysis of principles of taxation and public policies is undertaken.

In the same proportions and in approximately the same volume as public expenditures, tax collections have increased from decade by decade. And like expenditures, only after major wars has there been any significant reduction in taxes. "The tax burden" has become one of the major problems of modern life. Perhaps the principal cause of the universal discontent over taxation is a widespread impression that tax moneys, when paid over to the government, somehow disappear. This is not so, of course, for payments to governments reappear as government expenditures. It is the absence of the individual *quid pro quo* relationship in public finance that makes it possible for people to lament the increasing tax burden while demanding additional services from government.

The burden of taxation upon the individual assumes two forms. First, there are taxes which are directly paid to government, such as property, income, and automobile license taxes. Second, there are the taxes which

raise the prices paid by individuals or reduce the incomes they receive (if they are owners of resources in any degree specialized to the production of a taxed item). The tax on cigarettes, for example, tends to reduce the price received by tobacco farmers as well as to increase the price paid by the consumers of cigarettes. The first form directly confiscates a portion of the money income of the individual, leaving him with a smaller "income after taxes," or a "disposable" income smaller than his total income. The second type of tax does not so directly affect the money income of the individual, but by raising prices, it lowers his real income.[1] The issue of how real incomes are lowered by taxes is too little understood by taxpayers, who are generally not conscious of the entire weight of the tax burden falling upon them.

Public finance thus displays on both its revenue and the expenditure sides a curious phenomenon. Many persons do not realize the extent to which they are affected by some public services; e.g., their food may be cheaper or dearer because of agricultural controls of government. Similarly, they are perhaps only dimly aware that they consume public services more or less directly; the food they buy may be free from contamination because of the enforcement of public health standards. Moreover, when individuals or organized groups become actively interested in public expenditures, it is usually in a particular expenditure, such as better roads, or higher salaries for school teachers, or an expanded public works or farm program. Again the proponents of these measures are not fully aware of their tax burden.

There are reasons for this phenomenon other than apathy among taxpayers. For as a matter of fact, it is very difficult to calculate the entire tax burden falling on an individual. Chiefly, effects of taxes on prices cannot be exactly known. Suppose there were a tax imposed on oranges but not on grapefruit. The price of oranges would then rise; in consequence part of the demand for citrus fruit would shift to grapefruit, causing its price also to rise (though probably not so much as the price of oranges). The consumer of grapefruit has therefore been adversely affected by a tax on oranges. But it would be difficult to measure this, because the prices of both oranges and grapefruit would probably be affected by many other influences in the period in which such a tax would be in effect.

Several attempts have been made to measure the total tax burden on persons in given income groups. The results of one study are shown in Table 65.

[1] In Chap. 4 the relation of "indirect business taxes" to the measurement of the national output was discussed.

Table 65

TAX BURDEN BY INCOME GROUPS, 1938–39

Income class	*Taxes as percentage of income*
Under $500	21.9
$500 to $1,000	18.0
$1,000 to $1,500	17.3
$1,500 to $2,000	17.8
$2,000 to $3,000	17.5
$3,000 to $5,000	17.6
$5,000 to $10,000	17.9
$10,000 to $15,000	25.5
$15,000 to $20,000	31.7
$20,000 and over	37.8
Total	**20.2**

Source: Gerhard Colm and Helen Tarasov, *Who Pays the Taxes?*, Temporary National Economic Committee, Monograph No. 3, Washington, D. C., 1940, p. 6.

Money incomes in the period 1938-39 were much lower than in the postwar period, but so were tax collections. In computing these figures, federal, state, and local taxes were taken into account as they ultimately affect individuals' incomes. It is significant that persons with incomes of less than $500 should pay $120 in taxes, for it highlights the fact that the consumption of basic commodities is heavily taxed. Such essential articles as sugar are taxed, and many states have retail sales taxes. Tobacco is widely consumed and bears a heavy tax. Moreover, even a person with $500 income has some personal property or real estate (especially farmers) which is taxable. The figures indicate that the lowest income groups bear a heavier burden than the middle-income groups, though not more than high-income groups (who are more drastically affected by income taxes).[2]

Taxes may be conveniently divided into three major groups—personal, business, and consumption taxes. The important personal taxes are those on the incomes of individuals and estate and inheritance taxes. Business taxes comprise taxes on corporation net income, the social security taxes on employers, and a series of other levies on business enterprise. The con-

[2] For the postwar period, 1945 on, all the ratios are somewhat higher, because the increase of incomes is offset by the larger relative rise in taxes. In the higher income brackets the ratios are greater too, because the income tax occupies a much larger place in the tax system than it did in 1938-39.

sumption taxes are those imposed upon commodities and services, and include retail sales taxes as well as the property tax on nonbusiness property.

THE FEDERAL TAX SYSTEM

The taxes levied by the Federal government can be grouped in the following classes:

1. Individual income tax
2. Corporation income taxes
3. Estate and gift taxes
4. Social security taxes
5. Consumption taxes

During and in the period immediately after World War II, the bulk of all federal revenue was obtained from the individual and corporation income taxes and from consumption taxes (e.g., the cigarette and liquor taxes). This has not, however, always been the case. The tax revenues of the Federal government have passed through several distinct stages.

In the period before the War between the States, nearly all federal revenue was derived from customs duties—taxes on imported commodities—besides small amounts of revenue realized from the sale of land. In the first half of the nineteenth century, except for the effects upon prices caused by customs duties, the Federal government did not by its taxation influence production, savings, investment, and the distribution of income. During the War between the States a whole series of new taxes was levied, including an income tax and a large number of imposts on commodities and services. In the period following, the revenues of the Federal government consisted mainly of the income from liquor and tobacco taxes and from the customs duties, which items remained as the mainstay of the federal revenue system until the time of World War I.

The Sixteenth Amendment to the Constitution (1913) made it possible to levy an income tax and so enabled the Federal government to meet the financial demands of the war four years later. The income tax on individuals, on corporations, and an excess profits tax on corporations, together with a long list of commodity taxes, raised the tax revenue of the period during and just after the war. During the 1920's personal and corporation income taxes comprised half of the federal revenue, the remainder being derived principally from tobacco taxes and customs duties. Liquor taxes ceased to exist after the Eighteenth Amendment, and the commodity taxes initiated during World War I were abolished.

During the 1930's individual and corporation tax rates were increased, but the proportion of total federal revenue raised from these taxes was less than it had been in the 1920's. Partly this was because yields of income taxes during a depression are small, simply because incomes are low. Moreover, the adoption of several new taxes (additional commodity taxes, liquor taxes, social security tax, etc.) lowered the ratio of income tax revenue by increasing tax revenue from other sources.

During World War II, as well as in the period immediately following, federal revenues increased sharply. This can be seen in Table 66.

Table 66 FEDERAL TAX REVENUES

(in billions of dollars)

Fiscal year	*Total revenue*	*Individual income tax revenue*	*Corporation income tax revenue*	*Commodity and service taxes*
1940	5.7	1.0	1.1	2.0
1944	45.4	18.2	14.6	4.8
1945	47.7	19.0	11.9	6.3
1946	44.2	18.7	12.5	7.0

Source: United States Treasury Department.

These data show the increase in the major items of revenue. Other revenue sources not directly shown include the social security taxes, the estate and gift taxes, customs duties, and such nontax revenues as seigniorage.

The Individual Income Tax

The federal tax levied upon the incomes of individuals now dominates the tax system of the United States and in all likelihood will continue to do so. Yet the individual income tax is a relatively recent development. Before World War II, it yielded only about one sixth of total federal revenue, and only a fourth as much as the property tax of the state and local governments.

Except during the War between the States there was no income-tax law until 1894, when a measure was passed largely in response to popular sentiment against trusts and concentrated wealth. The income-tax statute of 1894 was, however, declared unconstitutional by the Supreme Court, on grounds pertaining to the provisions of the Constitution relating to direct taxes, and it seems also on the ground that the income tax as a tax was regarded as "objectionable" and "discriminatory." The income tax

was widely denounced at this time and previously as a radical—and vaguely socialist—device which threatened the existence of a free economy. In 1909 Congress provided for a tax on the net incomes of corporations under the name of an excise tax; and with the adoption of the Sixteenth Amendment in 1913, the income tax on individuals was constitutionally established.

The tax on the incomes of individuals is imposed upon incomes *as they are defined by law and by the regulations of the Bureau of Internal Revenue* of the Treasury. The definition of Income for tax purposes varies greatly from the ordinary conception of individual income.

In the first place, certain kinds of receipts are not included in taxable income, such as interest on state and municipal bonds. Constitutionally, the Federal government does not have the power to levy a tax on this type of income. A wealthy man could not, however, escape the effects of income taxation by selling all his assets to buy tax-free state and municipal bonds. Because the income from these bonds is tax free, the demand for them has grown with the rise in the federal income tax. As a result, the yields of state and municipal bonds are relatively quite low. The owner of these bonds pays no federal income tax, but his interest income could be about the same as if he owned taxable securities; if an individual's income is very large, however, he does gain by owning state and municipal bonds. A great advantage did accrue to persons who bought state and municipal bonds before the rise of the tax rates in the 1930's, and especially before the wartime increases of the early 1940's. The possibility of avoiding federal income taxation by holding tax-exempt securities probably results in a dearth of risk-taking capital because the very people who have enough capital to diversify their risks are the ones who gain most from buying tax-exempt bonds. Another type of receipt which is not income for tax purposes is gifts. If an individual receives a regular sum every month as a gift, it is not subject to income tax. The donor, however, is liable for payment of a gift tax, providing the gift exceeds a certain amount.

In the second place, there are deductions from gross income. These include interest, certain taxes paid, business and professional expenses, and —subject to limiting provisions—medical expenses and contributions to charitable and religious organizations. An attempt is made to distinguish between those expenses which are part of the cost of earning an income and those expenses which are a part of final consumption. However, as the mixed classification of these deductions would seem to indicate, there are borderline cases where the line drawn by law and regulation between deductible and nondeductible expenses is necessarily an arbitrary one. For example,

the interest deduction gives an advantage to those persons paying for houses and other durable consumer goods on installment plans. Some deductions are allowed, of course, to simplify administration of collections and the taxpayer's preparation of his return; the standard deduction allowed all persons in the lower income group (the "short form" return) is an example.

Similar to deductions, personal exemptions and credits for dependents are also allowed. In 1947, a $500 deduction was allowed per person in the family of the taxpayer; formerly, it was much higher, as in the middle 1930's, when $1,500 for a single person and $2,500 for man and wife, plus $300 for each dependent child, were allowed. During the war servicemen received an additional $1,500 exemption. The personal exemption serves several purposes. It makes unnecessary the payment of a tax by those with the lowest incomes and is thereby a means of adjusting revenue to revenue requirements. Lowering the exemption had an effect similar to an increase of the rates. It is chiefly a device for simplifying administration, since it results in fewer tax returns. Moreover, the exemption makes it possible for Congress to recognize the standard-of-living concept. When the exemption was higher, as in the 1920's and 1930's, it was thought that the Federal government refrained from taxing the minimum income required for an acceptable standard of living. However, inasmuch as the exemption was steadily lowered as the cost of living went up in World War II it is questionable that any direct relation has been applied in order to maintain for low-income persons a certain standard of living.

The rates of the individual income tax are progressively higher as the taxable income increases. During most of its existence, the federal income tax has applied two separate rates to taxable income (income defined as such by law, minus deductions and applicable personal exemptions). These two rates, the "normal tax" and the "surtax," have been defined variously at different times. Generally speaking, the distinction between them has been that the "normal" rate is the basic low rate applicable to all incomes, while the surtax rate is specifically applicable to the higher incomes. During and after World War II, the surtax rates were applied, however, to even the lowest net taxable incomes.

During the war years the income-tax rates, which are the normal and surtax rates combined, began at a minimum of about 20 per cent of net taxable income and progressed to a maximum of 91 per cent. In contrast, the tax law of 1928 provided for a maximum rate of 20 per cent. It seems probable that income-tax rates will be lowered after the immediate problems of postwar readjustment have been dealt with. The tax rates applying to net

taxable income give an exaggerated impression of the severity of the tax. Net taxable income is smaller than income as it is ordinarily measured.

Until World War II, the income tax was payable on March 15 for taxable income received in the preceding calendar year. During the war a withholding—or collection-at-source—system was adopted. In this system of administration, the employer deducts the income tax estimated to be due from the employee's wage or salary and forwards the amount to the Treasury. The employee can make adjustments (such as large deductions or additional income) by paying additional amounts or by asking for a refund in a return he files each March 15. Those taxpayers whose taxes are not withheld are required to estimate their taxes in advance and to make payments quarterly, with final adjustment made in a return on March 15 for the income of the preceding year. The advantages of both withholding the income tax at source and of requiring current (estimated income) payments are considerable for the Treasury: administration is simplified, evasion is reduced, and changes in revenue follow changes in tax rates and income more closely; for most taxpayers, there is less of a subjective burden because the tax is paid in smaller installments and in the same year in which the income is received. For some taxpayers, the withholding method may make payment of the income tax almost painless, for the taxpayer never lays his hands on that part of his income which is remitted by his employer to the Treasury.

The income tax on individuals presents many problems, both of administration and of policy. One of these is the treatment of capital gains and losses, which has varied often since the adoption of the income-tax law. A capital gain (or loss) is the profit (or loss) realized when a taxpayer sells a capital asset, such as securities, or real estate, or machinery and equipment. The capital gain (or loss) must always be "realized" in order to be taxed; i.e., the asset must actually be sold. Mere appreciation in value of an asset which remains in its owner's hands is not taxable. In recent years Congress has provided that if a capital gain is realized in six months or less, it is taxable like other income. For example, if someone buys stocks, sells them within six months or less, and makes a profit, the profit is taxable directly with his other income. Long-term (more than six months) capital gains, on the other hand, are taxable at lower rates. Capital losses, on the other hand, can be deducted only within narrowly prescribed limits. The capital-gains provisions of the income-tax law have been subject to widespread and some justifiable criticism. To be taxed, capital gains must be realized, i.e., a sale must take place. Property which increases in value without being sold may be just as suitable an object of taxation. To tax

capital gains, but to limit the deductibility from other income of capital losses, means that the Treasury shares in business successes, but refuses to participate (through lower revenue) in business losses; business incentives may thus be weakened.

The individual income tax has, furthermore, presented many opportunities for avoidance of payments. The exemption of interest on state and municipal bonds has been noted. Some of the federal bond issues are partly exempt from federal income tax. Income-tax payments can also be avoided through use of the corporation. Suppose a man operates a business which is incorporated and in which he is the owner of nearly all the stock. Suppose that the profits of the business would put the owner in the 60 per cent tax bracket if all were paid out as dividends to the owner. Instead, however, the owner of the business retains the profits in the corporation, and draws from it (in salary) only as much income as is required to meet his consumption expenses. The profits are left in the corporation to be taxed at the corporation income-tax rate, which is lower than the higher individual income-tax rates.

Another device is the creation of trusts. By creating a number of trusts, e.g., for the benefit of his children, a wealthy person can divide his property and so split his total family income into several smaller amounts, thereby escaping the progressive feature of the income tax. He must, however, pay a gift tax on the funds put into the trusts. Similarly, taxpayers who reside in states having the "community of property" principle for husbands and wives have an advantage because family incomes can be divided equally between two taxpayers; the total of taxes is thus less than if one tax were levied on the entire income.

The Corporation Income Tax

The corporation tax is levied on corporate *net* incomes or profits, as these are defined by law. During the 1920's and the early 1930's, the rate of tax on corporate net incomes was about 12 per cent, fluctuating only slightly from year to year as Congress modified tax rates in the light of changing revenue requirements. From 1935 to 1942, corporation-tax policy was in a state of flux. A mildly progressive rate structure was adopted in 1935, a tax on undistributed profits was put into effect in 1936, and in the following years other substantial changes were made. During the war, corporation profits were taxed at the rate of 40 per cent, together with an additional tax on *excess* profits.

One of the most difficult problems presented by the tax on corporation

profits is its relation to the individual income tax. Since 1936 the earnings of corporations, if paid out to the stockholders, have been taxed twice. The corporation pays the corporation tax on its net income, and when net income is paid out to stockholders in dividends the stockholder pays the individual income tax on his share of the dividends. Although double taxation is very common throughout the tax system, its application to corporation earnings is not easily justified. It is often held that since other forms of property income are not doubly taxed, corporation earnings have been singled out for discriminatory treatment. On the other hand, the possibility of retaining earnings in a corporation permits the partial avoidance of personal income taxation if a corporation is owned by one person or by a few men working closely together. In addition to the tax on net income, Congress imposed in 1936 a tax on the retained or undistributed profits of corporations in an effort to cope with this problem. This tax on undistributed profits proved to be very unpopular, however, and was abandoned. But the problem of integrating personal and corporation income taxation still remains.

The wartime excess-profits tax on corporations was first levied in 1940; in the fiscal years 1943 to 1945 it brought in far more revenue than the tax on the net incomes of corporations. The excess-profits tax was levied in addition to the net income tax, being imposed on all profits in excess of a base amount. Corporations were given the choice of selecting a base amount from two figures: the average profit in the period 1936-39, or profits equal to a given percentage on the capital employed by a corporation.

The excess-profits tax was discontinued at the end of the calendar year 1945, for this kind of taxation is a wartime measure. Difficult problems of policy and certain inequities in application are apparently inherent in the excess-profits tax which would be intolerable in peacetime, at least in context of present-day tax philosophy.

The Estate and Gift Taxes

The Federal government imposes an estate tax upon estates of more than $40,000 left by a decedent. An inheritance tax, on the other hand, is levied on the separate shares of an estate, as the shares are transferred to the beneficiaries under a will. The federal estate tax on estates in excess of $40,000 is progressive, reaching to a rate of 77 per cent on the highest bracket. In order to obtain some degree of uniformity among the states, Congress has permitted the payment of the federal estate tax to be reduced, within certain limits, by amounts paid to state governments for state death (inheritance and estate) taxes. The effect of this provision is that

a large part of the revenue from death taxes goes to the state governments, and the federal revenue from the estate tax remains comparatively small. The federal gift tax supplements the estate tax. Gifts of large amounts are taxed at rates which are three quarters of the corresponding estate tax rates. If there were no gift tax, persons might avoid the estate tax by transferring property to heirs in anticipation of death.

Federal Consumption Taxes

Consumption taxes are those imposed on commodities or on services with the intention that the burden of the taxes shall rest on the consumer. Consumption taxes include import duties and excise taxes, which are imposed upon selected commodities, and sales taxes which are levied upon all commodities and services (with specified exemptions) sold, e.g., at retail. The Federal government does not levy a sales tax. Usually, the tax itself is paid to the government by a business organization such as a manufacturer. The burden of the tax rests on the consumer in the form of the higher price for the commodity or service. Federal excise taxes are imposed on such commodities as cigarettes, liquor, furs, jewelry, luggage, matches, playing cards, etc., and on services (e.g., travel).

The Federal government has always relied upon consumption taxes; import duties were at one time almost the sole source of revenue, and liquor and tobacco taxes have played prominent roles in the revenue system since the War between the States. The consumption taxes are easily administered, since the Treasury collects the revenue from manufacturers or from importers. Liquor and tobacco taxes are likely to remain a steady source of federal revenue, even if other commodity taxes are dropped (as happened after World War I), for popular moral sentiments support a measure which reduces consumption of liquor and tobacco. In fact, the liquor tax is regarded as part of the legal control over the distribution of alcohol. Tobacco taxes were adopted in the nineteenth century when the use of this commodity was more widely frowned upon than it is now; the mere fact that tobacco has been taxed for so many decades is enough to assure its almost certain continuance in the tax system. The other commodity taxes adopted during the 1930's and early 1940's have more of the air of emergency about them.

The excise taxes imposed by the Federal government fall generally upon commodities which are widely believed to be "luxuries." The purpose in selecting commodities on this basis has been to tax articles believed to be consumed by the well-to-do, and thus presumably items which the low-

income groups "should not" consume. But such products as cosmetics, tires and tubes, and radios are widely consumed by low-income as well as by high-income groups. Excise taxes, in fact, yield large revenues only if they are imposed upon commodities which are consumed in large quantities. The only importance of the "luxury" attribute is that it affords an easy means of justifying the selection of commodities for taxation.

The import duties and social security taxes yield revenue in connection with the exercise of other controls, namely, on the volume and kind of commodity imports, and in the operation of the social security system. Further consideration of these two problems is given in Chapters 36 and 40.

STATE AND LOCAL TAXES

The tax revenue system of the state and particularly of the local governments is dominated by the tax on property. Most local governments depend entirely on the property tax for their financial support. A general picture of the tax revenues of the state and local governments is given in Table 67.

Table 67 TAX REVENUES OF STATE AND LOCAL GOVERNMENTS, 1941

Source of tax revenue	*STATE*	*LOCAL*	*TOTAL*	*Per cent of total state and local tax revenue*
	(in millions of dollars)			
Property	268	4,224	4,492	49
Sales and gross receipts	1,033	98	1,131	12
Pay rolls	901	5	906	10
Corporation and business	586	181	767	8
Individual income	225	2	227	3
Inheritance, estate, and gift	118	1	119	1
Motor-vehicle fuels	913	7	920	10
Motor-vehicle licenses	433	24	457	5
Other	30	166	196	2

Source: Adapted from *Public Finance and Full Employment, op. cit.*, p. 108.

During the nineteenth century and until about 1920, the property tax was almost the sole source of revenue for the states as well as for the local governments. During the 1920's the extensive construction of highways led to the adoption of state gasoline and motor vehicle taxes. The relief and social security programs of the 1930's led in turn to the imposition of state retail sales and pay-roll taxes. Consequently, with the increase of state revenue from other tax sources, the general tendency for the states in the

1920's and 1930's was to relinquish property taxation, leaving this source of revenue to the municipal and other local governments. Property tax revenues did not, however, decrease; they continued to increase (as did state and local expenditures) during this period.

The Property Tax

The property tax is an annual tax upon the value of property situated within the taxing jurisdiction. The cities, counties, school districts, and other local governmental units impose taxes on the basis of assessments on individual properties. The tax rate is usually expressed in mills (tenths of a cent) or in dollars per hundred of *assessed* value. The effective tax rate is probably close to about 2 per cent of the *market value* of property. Assessed value is almost always lower than market value.

From the standpoint of taxation, property is customarily divided into three main classes: (1) real estate, (2) tangible personal property (automobiles, furniture, livestock, and other movable physical property), and (3) intangible personal property (stocks, bonds, bank deposits, and other paper claims to wealth). Originally, the property tax was a *general* property tax levied on all property at a uniform rate. However, under a *classified property tax*, such as exists in many states, different rates of taxation are applied to the different classes of property; some classes of property may be exempt altogether. Property is divided into classes variously (sometimes simply into real estate and all other property) under a classified tax.

The central problem of property-tax administration is the method and practice of assessment. Since in many states the tax falls on all property, assessment means that all property must be valued for tax purposes. Because the process of valuation is assigned often to small and untrained staffs, inevitably some values placed on property are inaccurate. The standards of value employed in assessment are also open to dispute. In most states this standard is "fair market value," but interpretations (and even information) of the changing market values may vary enormously. Assessment of personal property is generally left up to the judgment—and conscience—of individuals.

Furthermore, much property escapes taxation, sometimes even real estate. It is often found in a community that parcels of real estate had been omitted from tax rolls for years. Personal property is usually listed by the owner, not an assessor, so that often intangible personal property is concealed. Because of this inevitable possibility, many states exempt intangibles from the property tax or tax them at a very low rate.

The owners of real estate bear a disproportionate share of the property-tax burden, for other forms of property are taxed at lower rates or are exempt. Even if rates are equal there is a tendency to assess values of other forms lower than real estate. Thus among the owners of different kinds of properties there are inequities, just as there are among owners of the same type of property.

State Business Taxes

Business enterprises are subject to the federal tax on net incomes if they are incorporated, and their owners must pay the federal individual income tax if their businesses are not incorporated. Business enterprises must also pay the property tax on real estate, equipment, and inventories to state and local governments. In addition, many states levy a series of other taxes on business organizations. Corporations pay taxes for the privilege of incorporation. Formerly, many states imposed special taxes on corporations directly as a form of business organization, but at the present time the tendency is to tax the net *incomes* of business organizations (and in some instances, gross incomes). State net income taxes resemble the federal corporation income tax though based on lower rates. One complication for state taxation of corporate net income is that only that part of the income of an out-of-state corporation which is earned within the state may be taxed.

Another form of business taxation, important particularly in the South, is the occupational tax on individual types of business and occupations, such as movie theaters, pool rooms, undertaking establishments, barbers, plumbers, etc. These taxes usually occur where political opinion favors state control in some measure over the privileges of the businessman. Businesses taxed are generally those involving some public interest of health, safety, or morals, although the tax may also be applied to businesses as a means of limiting entry into their field.

The physical properties of public utilities, including railroads, usually extend over the areas of several counties, and often over several states. Thus a special problem in taxing their properties arises for local governments. How can the value of railroad property in one county be assessed? State governments usually tax public utilities on the value of their properties within the state, and then either divide the proceeds among the local governments by means of grants, or retain the proceeds to finance state activities. Public utilities are of course subject to other taxes as a business, including a corporate net income tax and possibly one on gross receipts as well.

The taxation of banks and of insurance companies also presents special problems. The assets of both of these types of business consist largely of securities which as "intangible personal property" are exempt from taxes or are taxed only at very low rates. Insurance companies are therefore taxed instead on their gross premiums and pay a property tax only on their real estate. Both state banks and national banks are treated alike by state governments, being taxed by special levies on bank shares, or franchise taxes on banks' net incomes, besides the property tax on the real estate owned by banks.

Chain stores are taxed in a few states, more to hamper their development than to raise revenue. Since any tax system reflects the strengths and the purposes of the major political groups, the chain-store taxation obviously shows the influence of well-organized groups of independent retailers.

Motor-Vehicle Taxes

The motor-vehicle taxes and motor-fuel taxes are distinguished from other kinds of taxes by the segregation of their revenues from other funds for particular expenditure. Revenues from other taxes are not segregated, but go into a general fund from which all kinds of expenditures are made (an exception in some states is the use of liquor-tax revenues for such purposes as education and hospitals). But motor-vehicle and fuel taxes are used for highway expenditures. Highway users therefore pay for public roads, but not entirely, for other revenues are needed when highway expenditures exceed motor-vehicle and fuel tax revenues, as happened in the 1920's and 1930's. The gasoline tax, which is the most important of the motor-vehicle taxes, is perhaps the least unpopular of all taxes. Its costs of administration are low, and it yields a steady and dependable revenue—virtues commending it to state treasuries. Most motorists realize that the gasoline tax is used for highway construction and do not oppose the tax.

The gasoline tax paid by any one person over a period of time is a rough measure of his use of highways and thereby of his responsibility in their construction and maintenance costs. Moreover, the out-of-state motorist must in this way pay for the use of the highways of the states he visits. Even the more expensive construction required for and extra costs caused by the operation of heavy vehicles, or high-speed traffic by all vehicles, is to some degree supported by larger gasoline tax revenues. For heavy and fast-speeding vehicles both use more gasoline per mile than those of average weight and those operated at moderate speeds.

The tax is thus remarkably flexible in adapting payment to use and cost.

But there remain some inequities nevertheless. Heavy vehicles cause additional costs in highway construction and maintenance out of proportion to their higher gasoline tax payments per mile. The annual motor-vehicle license tax, which in some states is so low as to bring in only enough revenue to meet the costs of registration and of license tags, is widely used to supplement the gasoline tax in paying for highway costs which are not covered by the latter. These taxes are based on the weight or horsepower or value or wheel base of the vehicle—or a combination of factors—and thus are greater for larger vehicles. Still, the license tax does not adequately cover additional costs of large-weight traffic because this traffic is mainly composed of commercial vehicles which use the highways more often than private ones.

Commercial vehicles—common carrier and contract trucks, and buses—present a more difficult problem of highway taxation in the relation between this taxation and the taxes levied on railroads. The motor carriers and the railroads are competitive: both are regulated and both are taxed. The railroads have long claimed that the motor carriers' tax burden is a smaller one, because the trucks and buses use the public highways and do not contribute adequately to highway costs. As a result there has been pressure to impose special taxes on motor carriers. Some states do tax the gross receipts of bus companies, as well as impose a ton-mile tax. The ton-mile tax measures both weight and distance traveled and is thus a good method of making the heavy vehicles contribute to the costs of highways.

During the financial straits of the 1930's some states diverted gasoline and motor-vehicle license-tax revenue to nonhighway purposes because, while other revenues were diminishing, it remained relatively constant. Diversion of highway revenue to nonhighway purposes can be defended on the grounds that the motorist is as responsible for support of general government as the smoker or drinker or retail sales-tax payer. Most motor travel is performed on city streets which are largely financed from general revenue rather than highway taxes. On the other hand, a similar case can be made against full support of highway costs by motorists, since highways yield many benefits to the general public such as military security, encouragement of commerce, etc.

State Sales Taxes

There are many forms of sales taxes, most of which were developed after World War I. A *manufacturers' sales tax*, such as exists in Canada, is imposed on the receipts of manufacturers. A *turnover* or *transactions tax*

falls on every exchange of commodities and services. *Gross receipts* or *gross income taxes* are imposed on the total sales of specified industries or types of business.

The retail sales tax, i.e., on the sales of retail establishments, has now been adopted by about half the states. In the depression of the 1930's the revenues of the state and local governments fell off at the very time that there was an increase in their expenditure requirements, especially for relief and old-age assistance. Since property tax rates were already high, there was considerable resistance to further increases. Retail sales taxes were therefore put into effect, sometimes as "emergency" measures because they yield a large amount of revenue quickly. The state of domestic emergency of the 1930's has passed but the retail sales taxes remain.

Retail sales taxation differs from excise taxation in this way: all commodities or services except those specifically exempted are subject to a sales tax, but only specific commodities or services are subject to an excise tax. State retail sales taxes commonly provide for exemptions, such as sales of foodstuffs, sales to religious and charitable organizations, and sales of commodities already taxed (e.g., gasoline and liquor). Since the rate of taxation is usually 2 per cent, small individual sales require either the use of mill tokens or the somewhat inequitable system of dividing sales into brackets, with each sale within a bracket bearing the same tax.

Retail sales taxes have been criticized as unjust to persons with low incomes, but the ease with which they are administered and the large yield derived from them have so far prevailed against such criticism. Interstate commerce complicates the administration of a retail sales tax to some extent. Under the Constitution the states may not tax commodities moving in interstate commerce, but the Supreme Court has nevertheless upheld their right to levy a "use tax" on articles entering a state which would otherwise be subject to the sales tax.

Many of the states impose excise, or specific commodity, taxes. Gasoline, tobacco, and admissions have been mentioned. Liquor is another important object of taxation; some states operate retail liquor stores, the profits of which constitute in effect the tax on liquor.

Other State Taxes

Many of the states impose taxes on the incomes of individuals, similar to the federal tax in the definition of income and in the methods of administration but widely contrasted to it in rates. The rates of taxation in the states are generally in a range from 1 to 5 per cent. Because the per-

sonal exemptions are also higher, the effective rates in the state income taxes are still lower; the revenue derived is therefore comparatively small.

In those states in which natural resources (other than the soil) are important, special forms of taxation on the owners of natural resources have been developed. The principal reason is that taxation of natural resources as property, under the property tax, may lead to harmful practices in their use. To put an annual tax on the value of standing timber is to provide an incentive to cut the timber sooner than otherwise. About half of the states provide for special tax treatment of forests. Methods vary, but the general principle is to tax the yield from forests. A tax on timber as it is cut is neutral so far as the rate of cutting is concerned; i.e., such a tax does not provide an incentive for earlier cutting. Mines and petroleum resources likewise are subject to special taxation both because it is difficult to measure their values for property tax purposes and because of the conservation policy.

RELATIONS BETWEEN TAX SYSTEMS

Since the beginning of World War II, federal taxes have dominated the national tax system. A few decades earlier, federal revenues were secondary in importance; more than half of the tax revenues of the country were collected by the local governments. Table 68 shows federal, state and local revenues for calendar years at five-year intervals since 1920 and for 1943, the latest year for which comparable data are available.

Table 68 FEDERAL, STATE, AND LOCAL REVENUES FOR SELECTED YEARS

(in billions of dollars)

Year	Federal revenue	State revenue	Local revenue
1920	6.7	0.8	3.1
1925	3.8	1.4	4.8
1930	4.2	2.1	6.2
1935	3.8	2.8	5.8
1940	5.4	4.4	5.4
1943	22.3	4.8	5.7

Source: Adapted from *Public Finance and Full Employment, op. cit.*, p. 115.

In the nineteenth century federal revenues were derived from customs duties and the excise taxes on liquor and tobacco, while state and local revenues were derived from property and business taxes. Since then, the

increase of federal taxes has impinged upon what were formerly local sources of revenue; the expansion of taxes by both the Federal government and the state and local governments has led them to tax the same objects. Thus, for example, there are federal gasoline taxes and federal liquor taxes, as well as state income taxes and tobacco taxes.

In each major tax field—income, business, and consumption taxes—there is overlapping and conflict among the three levels of government. The only important example of absence of conflict is the property tax, which the Federal government does not impose. To some extent, of course, there is coordination between Federal and state governments. Inheritance taxes paid to the states are partially deductible in payment of the federal estate tax. The federal gasoline tax would probably be higher were it not for the state gasoline levies; yet it is doubtful whether there should be a federal gasoline tax. State income taxes must be low in view of the high effective rates of the federal tax. In part also, the seriousness of overlapping of taxes is lessened by the grants-in-aid which the Federal government makes to the states for a variety of purposes such as old-age assistance and highway construction.

Conflicts among states themselves are another feature of the American tax system. For example, several states have claimed the right to impose inheritance taxes on the same fortune. Where should a fleet of air-transport planes be taxed? In the state where the company is incorporated, or where the main base is, or where the planes happen to be on December 31? Should income be taxed on the basis of where the taxpayer lives, or where he earns the income? How should corporations which do business in several states be taxed? The possibilities of tax conflict among the states are apparently without limit.

CHAPTER 33

Tax and Fiscal Policy

The immediate purpose of taxes is revenue, but their effect is greater than this. The tax system as a whole directly affects the level of investment and employment; it influences the volumes and kinds of commodi-

ties and services produced, and the distribution of income. These general economic effects of taxation must be considered along with the revenues yielded by a tax system. To put it most simply, the problem of tax policy is how to raise the necessary revenue and at the same time to adjust the economic effects of taxation in the desired manner.

To say that the effects of taxation must be adjusted in the desired manner is really to avoid a major question: the effects desired by whom? It is possible that a dictatorship could have a wholly rational tax system, i.e., one which would yield revenue and would achieve precisely the economic effects intended by the dictatorship. Taxes are political decisions, and in a democracy, political decisions are characterized by conflicts of purposes and by resolution of the conflicts through compromise. Precisely because of political conflict, the nonrevenue aspects of taxation are important. It is not too much of an exaggeration to say that each organized group in the economy seeks to have governmental revenue raised by taxes on other groups.

REQUIREMENTS OF A TAX SYSTEM

It is not meant to imply that high principle is absent in the formulation of tax systems. Some principles, presently to be discussed, are accepted by nearly all individuals and groups. Much thought, indeed, has been given to the matter of the requirements of a good tax system. Adam Smith wrote in his *Wealth of Nations* that a good tax system should possess the characteristics of equity, certainty, convenience, and economy. That is to say, Smith believed that a tax system should distribute the burden of taxes fairly among the citizens, that each taxpayer's obligation should be made as certain and as little inconvenient as possible, and that costs of administering the tax system should be low. Other writers since 1776 have in general approved Smith's "canons of taxation," although there has been an inclination constantly to reinterpret the meaning of equity in taxation.

Besides having the qualities of equity, certainty and convenience to the taxpayer, and economy of administration, a tax system must also possess suitability for nonfiscal control, and it must harmonize with national policy on the maintenance of high levels of employment and national output. These requirements of a good tax system are not likely to be fully realized in practice, because of the possibility of conflict among the requirements themselves. Thus, the requirement of equity may clash with the requirement of nonfiscal control, just as the latter may clash with the policy of maintaining a high level of national output.

Before the two major policy problems of taxation—equity and the relation of tax revenues to national output—are considered, the nonfiscal control aspects of taxation can be mentioned briefly. Suitability for nonfiscal control means that a tax system is also used for control of production and consumption. The liquor taxes yield revenue, but they are also intended to reduce the consumption of alcohol. State chain-store taxes have the purpose of discouraging chain-store retailing. Import duties provide protection of domestic industries from foreign competition. Thus, many individual taxes have a specific control purpose, besides their function of bringing in revenue. The principal reason for the widespread use of taxes as control devices is constitutional: the Federal government has been able through its taxing power to exercise control where the use of other powers under the Constitution was impossible or dubious. Any tax can be administered so as to achieve nonrevenue purposes. For example, the rates of taxation in the federal corporation income tax are lower for small corporations; hence the tax is said to "encourage" small corporate business. There are almost always exemptions in the levying of a tax; exemptions can be manipulated so as to benefit particular classes of individuals or businesses.

The Problem of Equity in Taxation

The problem of equity in taxation is the problem of the distribution of the tax burden among individuals. Assuming that a given number of billions of dollars is to be raised, how can it be collected most equitably from the population? This is a question of principle. In the nineteenth and twentieth centuries, two main principles have been advanced as solutions to the problem of justice in taxation. These are the benefit principle and the ability-to-pay principle.

According to the benefit principle, individuals should contribute in taxes in proportion to the benefits received by them from government. The benefit principle presupposes, however, two things. One is that it is possible to measure the benefits derived by each individual from government, and the other is that all government services are calculable on an individual basis. Neither of these suppositions is completely wrong, but again, neither is valid enough to serve as a foundation for apportioning the tax burden. It is impossible, especially nowadays, to calculate the value of the government services consumed by an individual. Moreover, much of the activity of modern government is clearly in conflict with the benefit principle. In the present generation, governments make large expenditures for poor relief, for relief to the aged, for assistance to the unemployed, for benefits to

veterans and farmers and for services, such as public health and housing, intended mainly for the low-income groups. Application of the benefit principle would have low-income groups pay more in taxes, as individuals, than persons in the middle- and high-income brackets.

There are, however, limited applications of the benefit principle in the American tax system. One of these is in highway finance, which was discussed in the previous chapter. The taxes imposed by local governments are also applicable to the benefit principle. Local taxes are mainly, of course, the property tax. Many of the services provided by local governments can be thought of as benefiting the owners of property. But the benefit principle in local finance breaks down in the case of such activities as education. School systems are not financed solely by the beneficiaries.

The real importance of the benefit principle in taxation lies probably in its giving strength to proportional taxation. The tax system in effect in the nineteenth century embodied proportional rates throughout. It was then that the principle of taxation according to benefit predominated. Proportional taxation is neutral in its effects upon the distribution of incomes: it does not affect the inequality of incomes. Proportional taxation likewise does not strike at the volume of saving. Thus while proportional taxation prevailed, the tax system was favorable to the accumulation of capital. In the twentieth century, progressive taxation, in which the rate of taxation increases with the size of the taxable object, has come to the fore. Progressive taxation has grown along with the acceptance of the principle of ability to pay as the criterion according to which the tax burden should be distributed.

The ability-to-pay principle means that each individual should contribute in accordance with his ability to pay taxes, and without special regard for the government services enjoyed by him. In a general and loose fashion, the principle of ability to pay fits the main facts of modern public finance, especially with relation to the education and welfare activities of government and the existence of progressive income taxation. Ability to pay as a concept also harmonizes with the major political ideas of the present generation. That "the privileged" should be taxed so as to provide government benefits to "the underprivileged" is, to many people, apparently the main content of ability to pay.

But just what does "ability to pay" mean? How can an individual's ability to pay be measured? The usual answer is that income is the index of ability to pay, although the wealth and consumption expenditure of an individual have been suggested as preferable measures. The relation, however, between income and ability to pay can be interpreted as not being a

proportional relation. The progressive income tax, when interpreted as a realization of the ability-to-pay principle, is an application of the idea that, as income increases, ability to pay increases more than proportionately. This can be shown by reference to the principle of diminishing marginal utility. When applied to money income, this means that the larger one's income, the less is the utility of one dollar of income. Or alternatively, the larger an individual's income, the less important are the wants successively satisfied with the addition to income. The fact that, as income increases, a steadily higher proportion of income is saved, on the average, would follow logically from the principle of the diminishing marginal utility of money income, because the principle means that the satisfaction of consumption wants becomes less important as income rises.

Under the ability-to-pay principle—and if it is assumed that the marginal utility of money income diminishes very rapidly—a tax of, say, 10 per cent on the incomes of two persons with different incomes, would not be equitable. True, the man with the larger income would pay a greater tax, but the burden of a 10 per cent tax would not be so heavy for him as the burden of a 10 per cent tax for the man with the smaller income. For example, suppose the two incomes are $10,000 and $3,000. The hypothetical 10 per cent tax would deprive the one of $1,000 and the other of $300. The former would give up perhaps a fairly expensive summer vacation, while the latter would probably have to sacrifice clothing and might have to eat cheaper qualities of food. The sacrifices, then, are held to be not proportional; the utility sacrificed by the wealthier man is thought to be less than that given up by the less wealthy man. To make the subjective sacrifices the same in both cases, it is required that the $10,000 income be taxed at, say, 20 per cent, if the $3,000 income is to be taxed at 10 per cent.

The case for progressive taxation as the realization of the ability-to-pay principle has just been described. It is well to note carefully the nature of the argument. First, it has to do with subjective states of mind: the greater ability to pay of a person with the larger income is not due to the fact that there is more money there for the tax collector to seize, but to the state of mind which attaches less importance to one dollar the more dollars there are. Second, it must be supposed that individuals do not differ widely in temperament and taste, that the utility of dollars diminishes rapidly and in about the same way for all of them. Third, it must be assumed that the tax authorities know how rapidly the marginal utility of money diminishes; otherwise the scale of progression might be wrong and injustice might result. Now, the second and third of these assumptions are dubious. Individuals probably do differ widely, and neither the tax authori-

ties nor any one else actually knows, or can know, in a quantitative sense just what is the relation between individuals and the subjective satisfactions they get from different quantities of money.

When carefully considered, therefore, the ability-to-pay doctrine is not firmly founded as a justification of progressive income taxation. But ability to pay in its subjective meaning is not the sole ground for the progressive feature of income taxation. A progressive income tax can be justified on revenue grounds. Moreover, it can be defended as a means of lessening income inequality. As a political fact, progressive income taxation is an expression of the equalitarianism characteristic of the twentieth century.

As a principle of apportioning taxes, ability to pay is not limited to subjective considerations. Indeed, for many purposes of tax policy, justice in taxation means only the relation of taxes to income. In other words, ability to pay has the more modest meaning of adapting taxes to income. For example, it is commonly asserted that a retail sales tax is unjust, because it violates the ability-to-pay maxim. A retail sales tax, in fact, is frequently labeled "regressive." Strictly speaking, a regressive tax is one in which the rate of tax decreases as the size of the taxable object increases. In this sense a sales tax is not regressive; it is proportional. But the annual amount of sales tax paid is a higher percentage of smaller incomes than of larger incomes. There are two reasons for this. One is that a higher proportion of smaller incomes is spent at retail, and the other is that the proportion of saving to income is lower for smaller incomes. Hence, as a matter of proportionality, a retail sales tax, or any other consumption tax, bears more heavily on a taxpayer the smaller his income is. And if ability to pay means proportionality to income, such taxes are unjust.

Even from the standpoint of justice, however, there is something to be said for the consumption taxes. These taxes supplement the income tax and lessen the inequalities which attend it. One source of inequality in income tax administration is underreporting. Probably several billions of dollars of income a year is unreported for tax purposes. Careful administration can reduce this, but not to zero. If all revenue were raised from the income tax, there would be an injustice caused by differing standards of ethics: the honest would bear too large a share of the burden and the dishonest too little. But in the case of consumption taxes, evasion is more difficult for the individual, being limited to smuggling and bootlegging. Since most consumption taxes are collected from manufacturers and other business enterprises anyway, and since the burden on the individual is in the form of a higher price, large-scale evasion is out of the question. Hence, a tax system of both consumption and income taxes will result in smaller

inequalities among taxpayers than tax systems composed only of either consumption or income taxes. The consumption taxes have the further advantages to the taxpayer himself in that they are paid in very small installments, and that they are paid in anticipation of consumption. The person who goes to the movies once a week and pays an admissions tax of ten cents makes the payment just as he is (presumably) looking forward with pleasure to the movie. If on the other hand, there were no admissions tax, but instead an additional tax bill of five dollars a year, the subjective burden would no doubt be greater.

The tax system of the United States does not reflect, except incompletely, the application of the ability-to-pay principle. A large part of total government revenue is raised from taxes on consumption and from taxes on business. Why is this? The reasons are, first, that justice must often give way to other considerations in the gathering of tax revenues. Justice must most frequently defer to revenue. Suppose a state government needs 50 million dollars additional revenue and needs it quickly. A consumption tax can bring in this revenue within a few months, whereas an income tax may not yield the desired revenue for, say, a year and a half. Furthermore the revenue from a consumption tax is usually steady, while the yield from an income tax is subject to sharp fluctuations. In the second place, the burden of consumption taxes is not fully recognized by those who pay them. Hence, a government can raise more revenue with less friction and opposition from consumption levies than from an income tax. Third, taxation is a political fact. In the past, most taxes were imposed by those with political power upon those without political power. The American Revolution was given impetus by objections to taxation without representation; in European countries in the eighteenth century and earlier the ruling classes paid no taxes. The American Revolution, and the corresponding political changes in Britain and France, established a system under which taxes have been imposed by democratic legislatures. In the United States, the House of Representatives must, under the Constitution, initiate all tax legislation. But still there is a tendency for the groups in power to tax those with less power. The rise of progressive income taxation can be partly explained as the result of a political development in which the politicians, representing the many with low incomes, have brought it to pass that the few, with high incomes, have been taxed heavily. Such an explanation, it must be emphasized, is only partly true. Before World War II only about four million persons paid federal income taxes. During the war the number of income taxpayers increased by at least tenfold. Political pressure, therefore, is no longer in the direction of the many taxing the few, but rather it has become

a question of sharing the income-tax burden. Furthermore the costs of government have become so high that they cannot be met through taxation of the few (those with high incomes). As a consequence the many have been forced to tax themselves to pay governmental costs.

THE GENERAL EFFECTS OF TAXATION

The economic effects of taxation can be examined from two points of view of equal importance. First, there are the effects of taxes upon investment, savings, and consumption, as well as upon the distribution of income. Second, there are the effects of individual taxes upon the prices and production of particular commodities and services.

Consider now the first group of effects, i.e., those upon the determinants of the size of national output. In large part, these effects were discussed earlier, and need be treated here only very briefly. In Chapter 5[1] it was shown that while the bulk of tax revenue is withdrawn from the income stream it is again added to the income stream as governmental expenditures are made. In Chapter 31 transfer expenditures were described; it was shown how production by private enterprise and consumption by private individuals are affected by taxes and expenditures.

Investment is affected by taxation primarily through the impact of taxes on business incentives. The taxes which may influence incentives most are the personal income tax, especially that part of it which falls on capital gains, and high rates on the high taxable income brackets. There is some indication that the federal income taxes on individuals and corporations have had the tendency to discourage risky undertakings. But taxes are only one factor determining producers' decisions,[2] and consequently their effects should not be exaggerated. However, there seems to be little question that growth of new and small enterprises is slowed by corporate and personal income taxes. The major source of funds for expansion of such enterprises is retained earnings, since small enterprises cannot easily sell securities. Taxation of earnings needed for expansion necessarily slows the expansion of smaller enterprises, because they must give up earnings which might otherwise be retained and used for expansion.[3]

The volume of saving is affected primarily through the income taxes. Since the bulk of saving is done anyway by corporations and by individuals in the higher income groups, the income taxes are the main reducers of saving. The distribution of income among persons is influenced to some

[1] Pp. 123-24.
[2] See Chap. 5, pp. 111-19.
[3] See Chap. 7, pp. 163-65.

extent by taxes and by public expenditures. The income tax lops off large amounts from the highest incomes. A part of public expenditures goes to the lowest income groups: the indigent and the aged. Other public expenditures benefit the veteran and the farmer. Taxes and expenditures therefore bring about a transfer of income from the rich to some of the persons in the lower and lowest income groups. On the other hand, it must be noted that through the consumption taxes the lowest income groups yield part (in 1939 more relatively than the middle income groups) of their incomes to government, which makes some payments to rich persons, e.g., wealthy bondholders.

The Effects of Individual Taxes

Taxes affect prices and thus they influence the allocation of economic resources. This is the second group of effects of taxes. This problem is usually referred to as the "shifting" and "incidence" of taxation. Shifting refers to the process through which taxes affect prices so that the burden of a tax is thrown onto someone else. For example, the cigarette manufacturer has the legal obligation to pay the cigarette tax; the law places the tax on him, not on anyone else. But having control over his price he is able to charge more for cigarettes to the wholesaler. The latter in turn sells the cigarettes at a higher price to the retailer, who in his turn sells to the consumer at a higher price. The demand for cigarettes is inelastic, so that higher prices do not reduce sales much. The tax is *shifted* to the wholesaler, to the retailer, to the consumer. The consumer bears the tax burden. The *incidence* of the tax (i.e., the locus of its burden) is on the consumer.

There is a widespread popular impression that taxes, especially those on commodities, are always shifted to the consumer. Actually, this is much too simple a view to take. Consider the cigarette manufacturer again. He shifts the tax, but nevertheless is affected by it, since his sales must be somewhat smaller, because the effect of this tax is the approximate doubling of the prices to the consumer. His sales are probably smaller and his profits probably less even though he passes on the full amount of the tax. Moreover, to the extent that sales are less with the tax than without it, the consumption of tobacco in cigarettes is lessened and tobacco growers face a smaller demand than they might have without the tax. The amount of land devoted to tobacco is therefore affected by the tax, and the price of tobacco is probably somewhat less than it might have been if there were no tax. The effects of a tax are widely felt, even though it is shifted. In the final analysis it is found that the tax on cigarettes is shifted in large part

to the consumer, but to some extent it is also shifted back to the tobacco grower.

Whether or not a tax on a manufacturer, or on a retailer or other seller, is shifted depends on the operation of the price-forming processes for the commodity or service which is taxed. A commodity tax can be shifted only by means of changes in prices or by alterations in the other terms of exchange, such as quality. Taxes can be shifted "forward" or "backward." The cigarette tax is an example primarily of forward shifting, i.e., to the consumer. To the extent that the cigarette tax would cause a lower price to be paid to tobacco farmers, the cigarette tax would be shifted backward. The processing taxes imposed on certain agricultural commodities in the period 1933-36 were in general shifted forward to consumers. But the tax on pork products was shifted backward to the hog raisers; in this case the farmers received lower prices, while the prices paid by consumers were apparently not affected by the processing tax.

To be shifted, a tax must have an effect on either the demand for or the supply of a commodity or service. Since consumption taxes are levied on producers or sellers this means that supply must be affected by a tax. The effect on supply is through the increase in the producers' or sellers' costs as a result of the tax.

Consider tax shifting under conditions of pure competition. It will be remembered that the short-run supply schedule shows that the amounts offered increase with prices, and that the demand schedule shows that the amounts which would be bought decrease with higher prices. Because it is an addition to costs, a tax on a commodity raises the price at which each amount would be offered. Therefore, there is a new supply schedule and a new and higher short-run price. If demand is highly elastic, so much less can be sold at higher prices that suppliers may find that the full amount of the tax is not covered by the increase in price. If demand is highly inelastic consumers will continue to buy nearly the same amount at the higher price. Thus, the extent to which a tax is shifted depends in part upon elasticity of demand.

In the immediate market period, if buyers expect a new or increased tax to raise prices, they may increase their rates of purchase, in which case a rise in demand forces up the price of the taxed commodity for a time. In the short run there is no doubt that a tax on a commodity produced under conditions of pure competition will be partially or completely shifted, for in the short run any increase in costs influences supply and therefore price. In the long run, and assuming that no other force than the tax intervenes, the tax has its full effect upon supply: marginal producers at the old equi-

librium price are eliminated and supramarginal producers reduce output; the price rises to a new long-run equilibrium determined by the new long-run supply schedule. Both producers and consumers are thus affected by the tax. Depending upon the nature of the supply schedule, the change in equilibrium price is less than, equal to, or more than the amount of the tax.

One thing is certain. The shifting of a tax on a commodity produced and sold under conditions of pure competition takes place because of the effect of the tax on demand or supply or both, and not because the producers or sellers want the tax to be shifted. It is only under conditions of monopoly and of monopolistic competition that the producers' decisions have anything to do with the shifting of taxes to consumers. A monopolist may in fact decide not to shift a tax, but rather to bear the burden himself. This will be true where the tax is levied on the monopolist's profits, not on the commodity he produces, and where his profits are at a maximum. It is better for a monopolist to pay a profits tax out of maximum profits, rather than out of any smaller amount. If he shifted the tax to his consumers by raising his price, he would diminish his profits, assuming that he had been charging the profit-maximizing price, and would have smaller profits-after-taxes than if he bore the burden of the tax. On the other hand, a tax directly on the commodity produced by a monopolist is shifted, because such a tax strikes at his costs and affects his calculations as to profits. He will not, however, necessarily raise the price by the amount of the tax, for to do so would in most cases reduce his profits-after-taxes more than if he raises his price by an amount less than the tax. If, however, the monopolist has only a vague idea as to his market, and if he operates by rule of thumb, he will probably raise his price by the full amount of the tax. A monopolist whose price is regulated by public authority will almost always ask for—and get—permission to shift new taxes to the consumers.

Under conditions of monopolistic competition, the effects on prices of commodity taxes are even more complex. Monopolistic competition frequently shifts the sphere of competition from price to quality, service, and advertising. Thus the burden may be shifted to the consumer in the form of poorer quality and service. If a tax is not a large one, it may not be shifted at all, the manufacturer preferring to bear the tax rather than accept the risks or incur the costs of setting a new price or changing the quality of his product. In general, however, where price-setting is done by rule of thumb in a fairly stable context, taxes are simply added to prices. In practice, prices are often raised by more than the amount of the tax, particularly when the old price, plus the tax, gives a figure which is just below a multiple of five or ten. The price for convenience is then simply

raised to the next highest multiple. For example, a 10 per cent tax on a 90 cent item may result in a new price of $1, instead of 99 cents.

The effects of taxes on income, either personal or corporation, are different from the effects of commodity taxes. In most cases they are not shifted: the burden remains where the law imposes the tax. Shifting of the individual income tax would mean that everyone paying this tax would demand and get a tax-offsetting increase in income from his employer, or his customers, or his patients or clients. In a few isolated cases, this happened during the period of general labor shortages in the early 1940's, when the withholding of income taxes began. But it could not happen as a general rule. The personal income tax is a uniform and universal tax; it falls on everyone with a high enough income, and therefore it cannot reduce the number of income-seekers, although the volume of services may be affected. If there were an especially heavy additional income tax on, say, lawyers, the effect would be in time to reduce the number of lawyers, which in turn would raise the prices of legal services, thus shifting all or part of the hypothetical tax. Again, if one state imposed an especially heavy income tax, the effect would probably be to cause some people to move from the state. If businessmen moved and took their businesses with them, there would be at first some unemployment in the state imposing the tax, so that wage incomes there would fall; the shift would in this case be backward to wage earners. However, it may be expected that some wage earners will also move out of the state; this would tend to restore wage rates. Those employers and other income receivers subject to the tax whose resources are immobile or whose markets are located in the state will bear the burden of the tax.

The corporation net income tax, so far as it is tax on pure profits, cannot be shifted to consumers. This is the opinion of economists, although businessmen always insist that this tax is passed on to their customers. But consider: a corporate income tax necessarily falls only on some corporations, because not all corporations make pure profits. There are always some which lose money. And of these which do have profits, some make large profits and some small profits. If the corporation net income tax were to be passed on to consumers, it would follow that the most successful ones would raise their prices most, the less successful ones only a little, and the losing ones would not raise prices at all. It is obviously ridiculous to imagine that corporations' prices are set in this way: that the larger a corporation's profits, the higher its prices must therefore be. Corporation prices are determined in accordance with market positions. Moreover, it can be pointed out that a corporation net income tax would affect prices only if it

caused investment to be shifted to other uses. If corporations operated only in selected industries, and if noncorporate business forms operated in other industries, the effect of the tax on corporation net income would be to shift investment between industries. Rates of output and prices would then be affected, and the tax thereby shifted wholly or partly to consumers. Conceivably, a consequence of the tax on corporate profits is that more capital is put into farming and housing, fields in which the corporate form of organization is not important. On the other hand, it may be argued that the advantages (tax and otherwise) of the corporate form are so great that there would be an undue amount of capital allocated to producing goods in the industries characterized by the corporate form, if it were not for the redressing of the balance by the corporate income tax. It is probably safe to conclude, then, that this tax is not shifted in most cases. On the other hand, this conclusion must be modified in the light of the practices of corporations in recent years. Reserves for the corporate income tax are set up, and charges to such reserves have come to be regarded as costs. When management treats this tax as though it were a cost, price policies are likely to be affected and may lead to a partial shifting of the tax through higher prices.

Other business taxes, however, are mostly shifted. This is true of taxes which affect all business firms, such as the gross receipts taxes of some states. The incidence of the property tax presents a difficult problem. It is probably true that part of the burden of the tax is borne by property owners, and that the rest of the burden is shifted to others, such as tenants. The property tax on land is borne by the owner of the land at the time the tax was imposed, because the tax reduces the capitalized value of the land. Likewise, increases in the property tax on land are borne by the owner at the time of the imposition of the increase. The capitalized value of the land (its sales value) is reduced by the state's confiscation, through taxation, of a share of the rent. Subsequent land owners thus escape the burden of property taxes on land.

The property tax on improvements is borne by investors to the extent that the tax is no higher than taxes on alternative uses of capital. To the extent that the property tax in one jurisdiction, or on one given kind of transferable capital, is higher than that in other jurisdictions, or on other types of investments, the property tax is shifted to tenants or other users of the services of the type of capital goods involved.

The burden of the tax system as a whole thus depends on how taxes are shifted. It is to be remembered that indirect effects are particularly important. All prices are affected by taxation, not merely the prices of those

commodities and services which are taxed. All prices are influenced in two ways: through price interrelations,[4] and through the effects on incomes brought about by taxes.

TAX POLICY

Any tax system reflects a variety of public purposes, some of them conflicting. Revenue, equitable distribution of the tax burden, ease of administration, the loci of political power—these are purposes among which compromises must be made. Revenue conflicts in part with equity, and equity is often in conflict with ease of administration. The question of the relation of taxes and of public expenditures to the level of income and employment is one that has emerged relatively recently. Even before the 1930's this question was not completely ignored. There was a notion that taxes ought not to "hurt business." In any case, the sum total of taxes, except in war periods, was too small to have much effect on the level of national output. Hence there was not much reason to think about the problem.

Earlier in this chapter, reference was made to flexibility of total tax revenues as a requirement of a good tax system. Flexibility in this sense means that tax revenues are low in periods of depression and high in times of relatively full employment. Emphasis on flexibility has come with the increased attention given to fluctuations in national output, and to the role of government expenditures and taxes as they affect national output. Before the 1930's, a good tax system was supposed to have the virtue of a total revenue as stable as possible from year to year. The cigarette and gasoline taxes, for example, yield a very steady revenue. This is due to the wide use of these commodities, and to their low-income elasticity, which in turn is due to the small part they play in the budgets of most consumers. On the other hand, an income tax yields a very unstable revenue: high in periods of prosperity and low in depression years. Stability of revenue depends not only on the characteristics of individual taxes. It also is affected by the number of taxes in a tax system. The larger the number of taxes, the more stable revenue is likely to be.

In the late 1930's some economists began to urge that the structure be overhauled so that taxes would impinge more heavily on savings and would strike less seriously at consumption. This policy was advocated during a depression period; its purpose was to reduce the propensity to save, to increase the propensity to consume, and therefore to add to the aggregate demand for current output and to make tax revenues as a whole more

[4] See Chap. 16, pp. 374-76.

flexible. The effect of these proposals would have been to raise the income taxes and to lower the consumption taxes. It was believed that higher income taxes would strike directly at savings. Congress paid no particular attention to these suggestions. Even if it had, Congress would have had to weigh these considerations against all the others relevant for tax policy—revenue, equity, administration, and the demands of special groups. It was easier to arouse interest in the late 1930's in using government expenditures as a stimulative measure than to indicate the advisability of flexible tax revenues.

Wartime tax policy was dominated by revenue demands. Income, business, and consumption taxes were all increased. Besides revenue, however, Congress had to see to it that taxes did not interfere with incentives, for it was necessary to mobilize the civilian population for war production. This was one of the main reasons income taxation could not be raised too fast and too far. Taxes were placed on a larger number of "luxuries" in order partly to cut down "nonessential" consumption. In a way, the tax policy of the war period was the reverse of that which had been recommended during the late 1930's: consumption was to be reduced, and savings were to be seized through taxation for use in war financing.

In the immediate postwar period, federal tax policy was confused. The heavy wartime taxes continued to be appropriate during the period of inflationary boom; consumption taxes tended to hold demand in check, and the income tax yielded a heavy revenue. But the excess-profits tax on corporations was quickly abolished, partly for reasons of business incentives, and Congress began to modify the individual and corporation income taxes with the same purpose in mind.

It is not to be expected that the tax system will be dominated by any one idea, such as the stabilization of national output and income at high levels. The higher the total tax burden becomes, the stronger will be group pressures to ward off taxation of themselves and to throw the burden onto some other group. The centrifugal forces of political democracy will probably prevent the integration of the tax system in the service of any one idea or power group.

BALANCING THE FEDERAL BUDGET

Until the 1930's it was the tradition to balance public budgets on an annual basis. During and for a while after war periods the federal budget had been out of balance—the excess of expenditures increasing the national

debt. But apart from that, hardly anyone had ever doubted the wisdom of balancing the budget every year. Balancing the budget simply means that expenditures are covered by tax revenues and by such nontax revenues as fees and fines; in other words, that no borrowing is necessary.

In the case of state and local governments, unbalanced budgets mean that borrowing is undertaken usually to pay for public improvements of a permanent nature, such as highways and school buildings. When borrowing is done for such purposes, it is considered justified. The reason is that borrowing means additional taxation for the future, to pay interest and principal on the debt. But this matches the services of the permanent improvement, services which will be yielded for a long period of time.

With the beginning of the depression of the 1930's, federal tax revenues began running behind expenditures. The deficit had to be met by borrowing, i.e., by selling securities and hence adding to the federal debt. The depression did not decrease federal expenditures, on the contrary it had the effect of causing demands for larger expenditures, but it did cut down revenue. At first, in the early 1930's, the unbalanced federal budget was to most people a cause of alarm, and the growing federal debt was regarded as a dangerous thing. But later, and under the Roosevelt administration, there arose a tendency to make almost a virtue of the unbalanced budget. There thus began a controversy, which has not yet died down, as to the wisdom of deliberately unbalancing the federal budget in order to use the federal expenditures as a lever on the rest of the economy.[5] This is the controversy over deficit financing.

Deficit Financing

The deficit financing of the Federal government in the 1930's was not on a very large scale. The deficits ran from 1 to 3 billion dollars a year. From 1932 on, the deficits can be attributed to the emergency expenditures, for RFC loans, for agricultural benefits, for relief of the destitute, for public works, and, in 1936, for the veteran's bonus. The growth of a *policy* of deliberate deficit financing was very slow.[6] The Hoover administration tried to balance the budget. Even Roosevelt campaigned in 1932 for lower taxes, economy in expenditure, and a balanced budget. In 1933, however, the

[5] See Chap. 6, pp. 131-32 and 142-44, for a discussion of increases and decreases in government expenditures as they affect national output.

[6] See Arthur E. Burns and Donald S. Watson, *Government Spending and Economic Expansion*, American Council on Public Affairs, Washington, D. C., 1940, especially Chap. 3.

emergency had become so acute that drastic steps were felt necessary. These were the reduction in the weight of the gold dollar, the National Recovery Administration program to stabilize industry, the Agricultural Adjustment program to stabilize agriculture, large-scale expenditures for relief, and a public works program. It was believed in administration circles that the public works expenditures would "prime the pump" of recovery and stimulate expansion of the producer goods industries.

Now a policy of deficit finance makes sense only if there actually is a deficit, i.e., *there are expenditures coming from newly borrowed funds, and not from taxes.* Keynes had published in 1933 a pamphlet entitled *The Means to Prosperity,* in which he argued that large-scale "loan-expenditures" (i.e., expenditures from borrowed funds) would stimulate the economy. They would do this, he said, through increasing incomes by more than the amount of the loan-expenditure itself. This would be the effect of the multiplier, a concept which Keynes made widely known in a later publication.

A policy along these lines was not part of the early New Deal program. The New Deal emphasis was at first on particular kinds of expenditures as stimulants to the economy. These were public works and work relief outlays. It was not until about 1936 that the nature of deficit spending began to be widely realized. A theory and a policy of "compensatory spending" was then worked out. This was the notion that in time of depression the Federal government ought to engage in deficit financing, and that the additional expenditures would act to increase the national output. In periods of prosperity the Federal government ought to achieve a surplus of revenue, the surplus being used to retire the additional national debt incurred in the depression period. Hence the federal budget would be balanced "over the business cycle," serving as a stabilizing force to temper the severity of depression and mitigate the boom.

By 1938 two things had happened to strengthen the interest in deficit financing. One was the decline in national output and employment in 1937-38, and the realization of the obstinacy of the depression itself, and the other was the dissemination and popularization of the views of Keynes as they appeared in his *The General Theory of Employment, Interest, and Money,* published in 1936. Thus deficit financing was a fully developed idea by the time the war began.

What are the merits of the deficit financing policy? Can the federal budget be used to stabilize employment and income, and at high levels? The experience of the 1930's is not conclusive enough to permit easy answers to

these questions. For one thing, the policy never was actually in full force. There were indeed deficits, but the policies of both the President and the Congress were on a year-to-year basis. In each fiscal year, there was always the intention to balance the budget in the near future. Actually the budget was not balanced, but the policies pursued were based on the ever-present hope that the depression would quickly go away. Again, the deficit expenditures of the 1930's were probably not large enough to achieve substantial increases in national output. Moreover, there was much mistrust in business circles of the government's fiscal activities and programs. The attitude of the New Deal toward business was in general hostile, and not conducive to new investment.[7]

During the war the federal budget did dominate the entire economy. Federal expenditures and deficits were so great that total effective demand exceeded the value of total output at prewar prices. Hence the wartime controls and the wartime price rises. At the end of the war the federal budget began to shrink, but in 1947 federal expenditures were four times the prewar magnitude, and for the foreseeable future they will continue to be much larger than before the war. The mere size of the federal budget cannot but have fairly substantial effects on the rest of the economy.

The question of fiscal policy in the postwar period must be regarded as still open. The relations among total savings, investment, and consumption are so complex that no one can be sure as to just what will be the actual effects of given volumes of deficit expenditures by the Federal government. After the end of the war the economy proved capable of full employment—in fact, of an inflationary boom—despite a drastic contraction of federal spending. Hence, the experiences of the 1930's are of little help; both the tendencies in the economy and in federal finance were quite different after the war.

With the next depression, however, it is a certainty that the Federal government will intervene with deficit expenditures. Such intervention will probably come too late to be fully effective, but it will come simply because of the pressure to "do something." In this respect—the acceptance of the idea that the Federal government must act—the 1930's did emphatically mark a turning point. The opposition to deficit spending is much less than it was; and even some business groups have given it qualified approval. Should the next depression be a relatively mild one anyway, and if there is deficit spending, it will promptly be concluded that the policy is successful because events will have "proved" it. Similarly, the failure of the policy to end the

[7] *Ibid.*, Chaps. 4-7.

depression of the 1930's probably meant only that the depression was too deep for so mild a dose of the remedy applied. The Employment Act of 1946 provided for a Council of Economic Advisers to carry on continuous study of the behavior of the economy and to make recommendations to the President and Congress on economic policy. The emphasis is on those economic policies which relate to the level of employment and national output, to such policies, therefore, as deficit spending.

Government Borrowing

When faced with situations in which expenditures are larger than tax revenues, national governments have met deficits in different ways. Before the nineteenth century they often debased their coins. Sometimes they also paid their obligations with notes which circulated as money. The greenbacks issued during the War between the States were forced into circulation; they were noninterest-bearing debts. Apart from these practices of the past, however, governments meet their deficits by borrowing. This is done by selling securities of different types.

The securities sold by the Federal government may be divided into three main groups: the "nonmarketable public issues," "special issues," and "marketable public issues." Of the nonmarketable public issues, the savings bonds are the most important, and of these "series E" the most familiar. The savings bond has been given various popular names—defense bond, war bond, victory bond—and has been advertised heavily. The savings bond cannot be sold; it can only be redeemed. Hence it is not marketable. The special issues are put out in connection with the social security program and the various government trust funds such as the fund which provides for pensions of government employees.[8] These issues are not sold but are held by the Treasury itself as an acknowledgment of its indebtedness to the aged and other persons covered by government programs.

The marketable public issues are of several types. *Treasury bonds* are long-term securities, ranging from 10 to 20 years. These bonds are issued in large denominations (usually $1,000) and they now bear interest at rates of 2 per cent to 2½ per cent. There are many different blocks or issues of bonds outstanding, because they were sold by the Treasury at different periods in the past. These bonds are bought and sold in the bond market, after they are first sold by the Treasury. Then there are *Treasury notes,* which are medium-term securities, with a maturity of 3 to 5 years, and with interest

[8] See Chapter 40.

now at 1¼ per cent. *Certificates of indebtedness* usually run for one year; they bore interest during the war and immediately afterward at ⅞ of 1 per cent. *Treasury bills* run for 3 months, bear no interest, but are sold at a discount. That is, the Treasury offers its bills for sale, a bill being a piece of paper which has the promise of the Treasury to redeem it at face value, in 3 months. The price at which these bills were bought during and immediately after the war usually meant an effective rate of interest of ⅜ of 1 per cent; i.e., a $1,000 bill was bought by a bank for $996.25. Rates were increased in 1947.

In issuing several kinds of securities the Treasury seeks to accommodate every kind of potential lender. Some lenders want to lend their money only for short periods of time; they buy bills and certificates because they are more liquid. Other lenders are willing to make loans for long periods; they will prefer to buy bonds. The savings bonds are issued primarily for persons of small and modest means. These bonds have fixed redemption values—no one can lose on them. On the other hand, marketable securities are bought and sold at fluctuating prices in the bond market, so that there is some risk in holding them unless the owner wants to keep them until their maturity date.

It will have been noted that the rate of interest varies with the length of life of the security. This means that the larger the proportion of short-term securities, the lower are the total interest payments which the Treasury must make.

At its peak at the end of December, 1945, the composition of the federal debt in billions of dollars, was:

Treasury bills	17
Certificates of indebtedness	38
Treasury notes	23
Treasury bonds	120
Nonmarketable public issues	57
Special issues	20
Total interest-bearing debt	**276**[9]

It can be seen from these figures that the savings bonds, which are included in the nonmarketable public issues, are a small part of the total, about one fifth.

The ownership of federal securities at the same peak period of December, 1945, is shown by the following figures, also in billions of dollars:

[9] The figures do not quite add up to this total, because they have been rounded off.

Commercial banks	91
Federal reserve banks	24
Insurance companies	24
Mutual savings banks	11
Business corporations	29
State and local governments	6
United States government agencies and trust funds	27
Individuals	64

These data show that about a third of the national debt is held by commercial banks, and that public bodies own about a fifth of it.

The interest-bearing debts of the state and local governments are on a much smaller scale. State and local securities are generally long-term bonds, the most distinguishing feature of which is the fact that the interest on them is not subject to the federal income tax.

The National Debt

The term *national debt* is a common one, but it may be somewhat confusing. It means the *interest-bearing debt of the Federal government.* It does not mean the entire obligations-to-pay of the Federal government; these include gold certificates and currency, which are not part of the national debt. It does not mean the public debt, which is the sum total of all interest-bearing federal, state, and local debts. And it must be distinguished from the sum of public and private debt. The national debt is, to repeat, the interest-bearing debt of the Federal government.

The Federal government began its existence in debt, and has been out of debt in only one year, namely 1835-36. The Revolutionary War, the War of 1812, the Mexican War, the War between the States, each caused debt. After each war retirement began, but could not be completed except in the single instance of the period following the War of 1812. In the period before World War I the national debt was about 1 billion dollars. The debt rose to nearly 26 billion dollars in 1919. In the 1920's, retirement began once more. By 1930 the debt had fallen to 16 billion dollars. In June, 1940, it was 43 billion dollars, having grown during the depression decade by almost as much as during World War I. The increase during World War II was to nearly four times the size of the debt in the middle of the calendar year 1940. Table 69 shows changes in the federal debt for selected years since 1914.

Table 69 THE FEDERAL DEBT, FOR SELECTED YEARS, 1914–47

(in billions of dollars)

Fiscal year	*Federal debt*
1914	1.1
1919	25.5
1920	24.3
1925	20.5
1930	16.2
1935	28.7
1940	43.0
1941	49.0
1942	72.4
1943	136.7
1944	201.0
1945	258.7
1946	269.4
1947	256.9

Interest on the federal debt is about 5 billion dollars annually. The average interest rate is approximately 2 per cent. The amount of interest payments depends on the composition of the federal debt. The higher the proportion of short-term securities, the lower the interest burden. The interest payments are about five times what they were before the war. They are not much less than the entire amount collected by the Federal government in taxes before the war. Interest rates paid by the Treasury were much lower in the late 1930's and in the early 1940's than they had been. World War II was a "2 per cent war" in contrast to the more than 4 per cent which the Treasury had to pay on its borrowings during World War I.

In what sense is the national debt a burden? It has been said that by borrowing to cover so large a part of the financial cost of the war, those costs have been thrust onto the shoulders of future generations. This is only partly true. In no sense could the real costs—in material resources and human suffering—of the war be passed on at all, except for the consequences of resource depletion caused by the war. It is therefore a question of the financial costs. These could be passed on in full only if the national debt were owed to foreigners, i.e., if all or most federal securities were owned by persons living in other countries. In such a case, interest and principal payments would have to be made out of the country, and this would mean, ultimately, the export of commodities to pay the debt. Hence, future genera-

tions would have to produce goods for export without any return except the doubtful satisfaction of paying debts contracted by an earlier generation.

But the United States debt is not owned by foreigners, except to an utterly insignificant extent. The debt is and will continue to be held internally. If anyone should worry about the burden on his grandchildren, he should remember that the taxpaying grandchildren will pay interest to the bond-owning grandchildren. This does not dismiss the problem by any means, although it has been argued that since the debt is an internal one, it is scarcely a serious problem at all. The fact that "we owe it to one another" actually, however, offers little consolation.

The real burden of the national debt consists in the following:

1. The annual tax collection and payment of about 5 billion dollars in interest will adversely affect other federal activities, on the assumption that Congress will be willing to levy in any year only some given amount in taxes. Hence, the welfare, public works, agricultural, and perhaps the national defense programs, will suffer. It is not possible to say how valid the assumption is, but at the very least it is sure that interest on the national debt makes available a somewhat smaller amount for other federal purposes.
2. If over the next several years the ownership of the national debt should change so that the bulk of it would be concentrated in the hands of the wealthy few, the payments on the debt, coming as they do from consumption and income taxes on the many, would intensify the inequalities of income distribution.
3. Under some circumstances, the payment of the interest on the national debt would have a depressing effect on the employment and output. This would be the case if the taxes collected because of the debt reduced consumption, and if the interest payment went to those who would save them but not invest them. Hence, aggregate demand would be reduced. Furthermore, as was shown in Chapter 11, repayment of the securities held by commercial banks reduces the supply of money. This could, at times, also have a depressing effect.
4. The existence of so large a debt means that there is always available a mass of perfectly safe securities. Investors may be more inclined to choose the safety of government bonds rather than to make riskier security purchases. Hence, the national debt may be a deterrent to investment in private industry.

Of these aspects of the real burden of the federal debt, possibly the first is the most important. As the level of national output grows secularly,

because of population and productivity increases, the relative burden of the debt will steadily diminish. Furthermore, some repayment of principal is to be expected. In fact, the debt is already smaller than its peak just after the war. Hence, the interest payments will become smaller in total amount and will diminish relatively as productivity increases. Further, if the long-term trend of prices is upward, as seems to be the case, there is an additional reason for the burden to shrink in relative terms; it will be easier to collect the interest-paying taxes if the purchasing power of the dollar continues to fall over the long period.

The management of the national debt presents many technical problems. The debt consists of a mass of securities, some of which are maturing all the time. Those which mature are paid off and are replaced by others, unless the Treasury is actually reducing the total debt. Hence, even when the volume of the debt is stable, the Treasury is constantly borrowing to replace issues which are becoming due. The Treasury must arrange the maturities of the securities outstanding so that it will not have to face at any one time in the future too large a refinancing problem. Further, the fact that the debt is constantly turning over as replacement borrowing occurs means that the interest burden could increase. If the level of interest rates should rise, the Treasury would have to pay the higher rates on its new borrowing, and this would affect the average interest charge on the entire debt.

Repayment of the national debt can take place only very slowly, simply because it is so large. The American tradition has always been to retire the debt, and there is no reason to suppose that this tradition will not be followed in the future. It is probable, however, that the debt will increase in future depressions, as deficit spending appears. Perhaps these increases will more than offset such retirement as will be possible in periods of budgetary surplus.

The securities which constitute the national debt are assets to their owners, viz., banks, insurance companies, corporations, endowment funds, and individuals. This obvious fact is not to be overlooked. If it were somehow possible to retire all of the debt at once, a calamity would ensue, for the present owners of the debt would then have money instead of the federal securities on which they earn interest, and there would be no alternative securities to buy with the same safety. The banking system would be completely disrupted, since a major part of the banks' earnings is interest on government securities.

CONCLUSION

Tax and fiscal policy have changed drastically within the last generation. Tax policy formerly was concerned with the question of the equity of particular taxes. Now it has to do with the general economic effects of the tax system. Fiscal policy—the tax, expenditure, and debt policies, considered as a whole—is new and has become the major instrument of federal economic policy. Fiscal policy in fact has displaced monetary policy from the centers of interest and action. Both tax and fiscal policy have become of much more importance because of the impact of World War II, but had this war never occurred, tax and fiscal policy would have come to the forefront anyway because of the economic developments of the 1930's.

PART VII

INTERNATIONAL ECONOMICS

CHAPTER 34

International Trade

Trade among peoples of different and often widely separated regions is of ancient origin. The early Phoenicians were wide-ranging sea traders; and by the time that Athens ruled the Mediterranean, trade between regions was on a large scale and included some of the complicated features of modern commerce. Bills of exchange and a rudimentary international banking mechanism were used to meet the needs of merchants who ranged from the Black Sea to the Straits of Gibraltar. Rome developed an extensive and complex system of trade to maintain her empire, the organization of which was not only military but also commercial. Since the Renaissance—the Age of Discovery—international commerce has encompassed all parts of the world. The rise of sovereign states, each with its own currency, system of law, and customs, has given to international trade its essential character in the present era. Each state has become vitally concerned in the commerce conducted by its citizens, or by instrumentalities of the government, and thus imports and exports are regulated and directed to further the interest of the state. Some countries such as Great Britain and the Netherlands virtually live by foreign trade; to others, like the United States, it is less crucial but increasingly important; and for a few nations, notably Soviet Russia, foreign trade is of small importance. However, for all modern nations international trade influences in varying degrees the character of the productive system, the volume of output, and consumption habits of a people.

Consider consumption habits first. Every day the typical American family uses many items of foreign origin: Brazilian coffee, Cuban sugar, Canadian

pulp in newspapers, Bolivian and Malayan tin in canned goods, East Indies rubber, Russian manganese in steel hardware, etc. Unless foreigners produced these things, shipped them to their ports, and sold them to American importers, the daily consumption of Americans would be materially different. Even though the United States does supply part of its demands for some of these things, and even though undoubtedly more could be supplied if necessary, to cease importing would be a costly procedure and would reduce consumption. The importance of this trade to consumption habits, moreover, is not alone measured by a country's total volume of goods imported. The value of imported goods is normally only 2 to 3 per cent of the value of this country's national output. Quantitatively this amount appears insignificant, but were it to be discontinued its qualitative effect on the consumption of all the population would be substantial.

Productive activities and volume of output are also vitally conditioned by foreign trade. Over half the cotton and one third of the tobacco produced in this country are exported; approximately a third of the annual output of office appliances is shipped abroad, and between 10 and 20 per cent of the machine tools. Automobiles, tires and tubes, locomotives, and fabricated steel are among the important export items. In varying degrees American productive activity or productive capacity has been built up in expectation of foreign demand; a loss of exports would be disastrous to some segments and more or less damaging to others. Furthermore, the productive structure has gradually come to rely on imports of raw materials. The metallurgical industries, for example, require ore imports without which much of their productive capacity would lie idle.

Like the imports, the exports of the United States are not large in relation to national output. But both exports and imports are vital to the activities of certain industries which themselves exert a substantial influence on the character of the American economy and on the level of its national output. Similarly, American imports and exports influence the economic activities of foreign countries. The American market has long been of utmost importance to Brazil, Cuba, and the Philippines, and is at present the chief economic support of Europe. American machine tool exports will remain essential to the industrialization of Latin America and the modernization and expansion of British and European industrial economies.

Each country in the network of international trade is both a market and a source of supply of commodities; each contributes to the general level of economic activity of all others. This principle has been clearly demonstrated by the dependence of many European countries on imports from the United States following the defeat of Germany in World War II. Before the war,

Germany had supplied many of the European countries with steel, coal, machinery, and other commodities essential to their industrial activity; Germany in turn was an important market for the products of other European countries. The disruption of the German economy because of war and occupation policies necessitated the importation of huge amounts of American goods into these European countries. Indeed, the general decline of European productive activity and standards of living may be traced largely to the disruption of the German economy, for Germany had been the center of continental trade and industrial production. The political considerations which kept the German economy prostrate in the early postwar years increased the dependence of western European countries on American goods—and American relief. Whatever may be said of the wisdom of these political considerations, their economic consequences are evident enough: without trade among themselves and with the rest of the world, the European nations can live only in poverty.

As time goes on, the dependence of the United States on foreign trade will undoubtedly increase, although this trade is unlikely ever to be so important as is the trade of England and western Europe to those countries. Increasingly the United States relies upon foreign nations for raw materials to supply its enormous industrial capacity and for markets to absorb an increased amount of its industrial output. Saying this is no plea for greater "international collaboration," as current fashion phrases it. The greatest expansion of world trade occurred in a period when this phrase was yet uncoined. Foreign trade is a matter of buying and selling, and like all trade it is undertaken in expectation of benefits that all parties hope to get from the transactions. It is not from a sense of collaboration but from a hard-headed realization of economic (and, increasingly, political) advantages that the growth of trade may be expected in future years.

The following discussion examines the bases and principal features of world trade, the import-export position of the United States, and the interdependence of a nation's imports and exports. Chapter 35 analyzes the complex question of international finance, including the balance of payments, foreign exchange rates, exchange stabilization, international loans, and certain international financial institutions created in recent years. Chapter 36 deals with the policies of nations in the conduct of foreign trade and financial activities and the influence of foreign trade on the level of national output.

THE BASES OF INTERNATIONAL TRADE

International trade implies a specialization of productive effort; it both results from and brings about an international division of labor. In this respect international trade corresponds to trade among persons and groups within a country. When individuals specialize in practicing law, growing wheat or tobacco, selling gasoline, and in the innumerable other activities making up the nation's productive effort, they perforce rely on others to supply them with the variety of commodities and services they want. Regions within a country specialize along certain productive lines. The bulk of this nation's coal is mined in the Appalachian region from Pennsylvania to Tennessee; most of the cotton is grown in a half dozen of the southern states; steel comes largely from the Pittsburgh, Chicago, and Birmingham mills, tires and tubes from Ohio, lemons from California. The internal or domestic trade of the nation has its source in this individual and regional specialization. So it is with international trade. In 1938 India produced 99 per cent of the jute entering into world trade; southeast Asia produced 91 per cent of the rubber, 81 per cent of the coconut oil, and 50 per cent of the tin. Most of the world's coffee comes from Brazil and Colombia, tea from Ceylon, automobiles and business calculating machines from the United States, newsprint from Canada, farm machinery from England and the United States, olive oil from several Mediterranean countries.

Specialization of productive effort means that a country produces more of some things than it requires for its own use. The United States normally produces more cotton, office equipment, tractors, tobacco, machine tools, and many other things than is demanded for its own consumption. Other nations which produce none of these things, or an insufficient amount of them, demand and buy this surplus production. They in turn produce more coffee, rubber, manganese, lead, olive oil, than they want for their own purposes and sell these things to the United States. In effect, therefore, cotton is exchanged for tin, automobiles for olive oil, machine tools for coconut oil, tobacco for rubber, tractors for lead, etc. In practice, of course, these exchanges are not made on a commodity-for-commodity basis. As a rule, each country sells its surpluses abroad in the most favorable markets, and buys what it wants from the most favorable source of supply. The point of importance is the specialization of productive activity.

This has two results for each country: (1) a surplus of the things in which specialization is marked, and (2) a deficiency in other things which are not produced at all or in amounts inadequate to meet demands. The economic

resources of the political (or geographic) subdivisions of the world produce too much or too little of many goods to meet their individual demands. Thus the patterns of resource use are not appropriate to the patterns of wants in each country; international trade corrects these disparities which are inherent in specialization. In a nation like Great Britain these disparities between production and consumption are great and must be equalized by a large volume of both imports and exports; on the other hand, a nation like Soviet Russia, which produces little that it cannot use and consumes most of what it produces, is not in need of extensive trade to balance its economy.

The underlying causes of international specialization of productive effort and the international trade that develops from it must be examined. Why does India produce most of the world's jute, the United States most of the mechanical refrigerators? Why does virtually every country specialize in some important respects?

Nature has a hand in part of this specialization. Tropical fruits can only be grown economically in the tropics; wheat, cotton, corn, tobacco, and vegetables require moderate climate. Iron ore, petroleum, lead, tin, copper, and other minerals essential to modern economic life are also distributed unevenly. The same can be said of forest resources. Climate, soil, minerals, forests, and physical accessibility thus show an infinite variety throughout the world and impose upon each country its own peculiar pattern of productive activities.

Apart from the natural environment, man himself accentuates this diversity of productive activity. Capital resources are concentrated. Before World War II, about four fifths of the world's capital facilities were to be found in the United States, Great Britain, and Germany. Human skills and aptitudes also vary greatly. Watchmakers in Switzerland, rug weavers in the Middle East, china craftsmen of Dresden, diamond cutters in Amsterdam, and many other highly specialized groups whose skills have become traditional over the generations contribute to the concentration and therefore diversification of productive effort.

By and large a nation, like an individual, tends to specialize in the production of things for which its combined resources are well suited: the production of those goods which best satisfy wants is maximized by concentrating economic resources on certain types of production and importing other goods less suited to a nation's resources. Note the statement that a nation *tends* to act so, for there are exceptions to this process which will be discussed later. The general validity of the proposition, however, is supported by much evidence. The raw-material- and food-exporting countries

such as Latin America and southeast Asia undoubtedly utilize their natural resources, labor, and capital better by growing coffee, wheat, raising cattle, mining tin, and producing rubber than they would if they attempted also to produce radios, automobiles, tires, typewriters, calculating machines, and other complex industrial products. Although eventually these countries may develop resources suitable for the production of industrial products in large amounts, for the present they do better by concentrating on raw materials and food products. An industrial country, such as Great Britain, would suffer a decline in national output by shifting from industrial output to an increased production of food because the present state of her resources favor (and are committed to) the former. However, the tendency stated here is partly offset by the endeavor of nations to gain a greater measure of economic self-sufficiency either for military security or for various economic purposes.

It may be asked: What is meant by "concentrating in the fields of special advantage"? This phrase might mean (1) specialization on those things in which a country is more efficient than any other country, or (2) specialization on those things for which its resources are best adapted. The difference is important.

Consider the first point. Suppose in the United States anything could be produced with smaller quantities of resources than anywhere else in the world. The United States would then have an *absolute* advantage in the production of all commodities, i.e., the output per man per day of *all* things would be greater than in other countries, or again, this country would produce everything at a unit cost below the foreign unit cost. Under this wholly imaginary circumstance, it might appear that the United States would import nothing from abroad. However, even under this hypothesis the United States would find it advantageous to import goods from abroad, *if within this country some goods could be produced relatively more economically than others.*

This brings the second consideration: specialization on those things for which a nation's resources are best adapted. Even if this country produced per man more of all products—or produced everything at a lower unit cost—than the rest of the world, there would be particular commodities produced within this country at relatively lower costs than other commodities. Thus the production of some things has a *comparative advantage* over the production of other things. This concept of comparative advantage does not rest upon a comparison of productive output per man—or unit costs of production—between two countries; it is solely a comparison of the produc-

tivity of labor, capital, and other resources within a country. Suppose all products are grouped in two classes A and B; and the output per man per day of product A is worth $50, while the output of B is worth $25; and the national labor force is composed of 100 men. Would this country produce both A and B? It would not if it sought to maximize production from its labor and other resources. By concentrating on A, it would produce a larger national output than by producing both A and B. If the labor force produced A, the value of total output per day would be $5,000 ($100 \times \$50$); if A and B were both produced—say, 75 men produced A, and 25 men produced B—the total product would be worth $4,375. Clearly it would be advantageous to concentrate on A, and to export part of the surplus of A to buy B from abroad.

The principle of comparative advantage is found at work whenever specialization is undertaken. If an executive of a large air line is also a competent flyer, would he be likely to divide his time between managing the company and flying its passenger planes? Would it not be more profitable for the company if he attended to the complex problems of management and employed a pilot to make the flights? Illustrations of this type could be multiplied, and all would assert the principle of comparative advantage: maximum productivity in a given area or occupation is secured if resources are employed in the work for which they were best suited, and if alternative and less productive work is assigned to others.

Similarly, in each nation the industries that loom large in productive activities and supply products for export are generally those which are *relatively* most productive. The imports of each nation tend to be those which (1) cannot be produced at all (for example a mineral not found in the country) or (2) can be produced, but at relatively high cost (or low productivity) when compared with the export industries. A classic statement of the matter is given by Adam Smith:

> It is a maxim of every prudent master of a family, never to attempt to make at home what it will cost him more to make than to buy. The taylor does not attempt to make his own shoes, but buys them from the shoemaker. The shoemaker does not attempt to make his own clothes, but employs a taylor. . . . What is prudence in the conduct of every private family, can scarce be folly in that of a great kingdom. If a foreign country can supply us with a commodity cheaper than we ourselves can make it, better buy it of them with some part of the produce of our own industry, employed in a way in which we have some advantage. The general industry of the country . . . will not thereby be diminished, no more than that of the above-mentioned artificers; but only left to find out the way in which it can be employed with the greatest advantage. It is certainly not employed to the

greatest advantage when it is . . . directed towards an object which it can buy cheaper than it can make. . . .[1]

A cheaper commodity, Smith held, was one which could "have been purchased with only part of the commodities, or, what is the same thing, with a part only of the price of the commodities, which the industry (human effort) employed by an equal capital would have produced at home, had it been left to follow its natural course." If the productive activities of the nations of the world followed this principle closely, then: (1) the economic resources of each would be employed in the most important uses; (2) for any level of employment, the value of the national output of each country would be maximized; (3) the value of world trade would be maximized, for each country would require more imports and have larger surpluses for export. These points are inherent in the optimum international division of labor. Such a state of affairs is possible only under conditions of free trade throughout the world. But since free trade does not prevail, the advantages accruing from specialization are not so large as they might be. These losses, however, may be offset to greater or lesser degree by other advantages that accompany economic diversification. Such matters will be discussed later. Despite the obstacles raised against the fullest development of international specialization, the great volume of world trade testifies to its importance.

Yet there remains the issue of how international specialization comes about. To say that "England" specializes in machinery, that "Argentina" specializes in beef, and so on for nations throughout the world is to obscure the historical fact that the active agents in international specialization are individual and organized business enterprises and not the national governments themselves. The use of these national names is largely a matter of convenience to designate the national origin of the enterprisers. To supply raw materials for the industrial nations, the great raw-material-producing areas were developed by industrial business interests. Discoveries of new lands as well as immigration to them have in large part been impelled by enterprisers in search of markets for both buying and, eventually, selling. The tea of Ceylon, the cotton of India, the rubber of the Netherlands East Indies, the tin of Malaya, the wool of Australia, the wheat and cattle of Argentina, Canada, and the United States, the coffee of Brazil—all were developed and promoted first by the British and Dutch and later by American, German, and French business interests. These individual and corporate enterprises in turn released upon the world an army of salesmen to sell the specialized manufactured products created in part from the raw materials

[1] *Wealth of Nations*, Cannan edition, Modern Library, p. 424.

and with the food of these outlying regions. The criterion of success for these ventures was the price-cost relationships. Raw-material producers had to compete in world markets as did industrial producers of finished products. The pressure of competition impelled specialization as the means of keeping costs and prices down. In the present-day world this continues, but it is modified by many factors such as cartel arrangements, governmental intervention, and the change from raw material to industrial activity in certain areas. The intense rivalry of individual producers in the international field has been modified—and in some respects aggravated—by the growth of political collectivism in the present age. Collectivism inevitably entails governmental control of private trade or at least governmental participation in trade, and in some cases even the suppression of private trade. As a political force it alters the traditional pattern of world commerce and transforms the conditions based on the principle of comparative advantage.

THE NETWORK OF WORLD TRADE

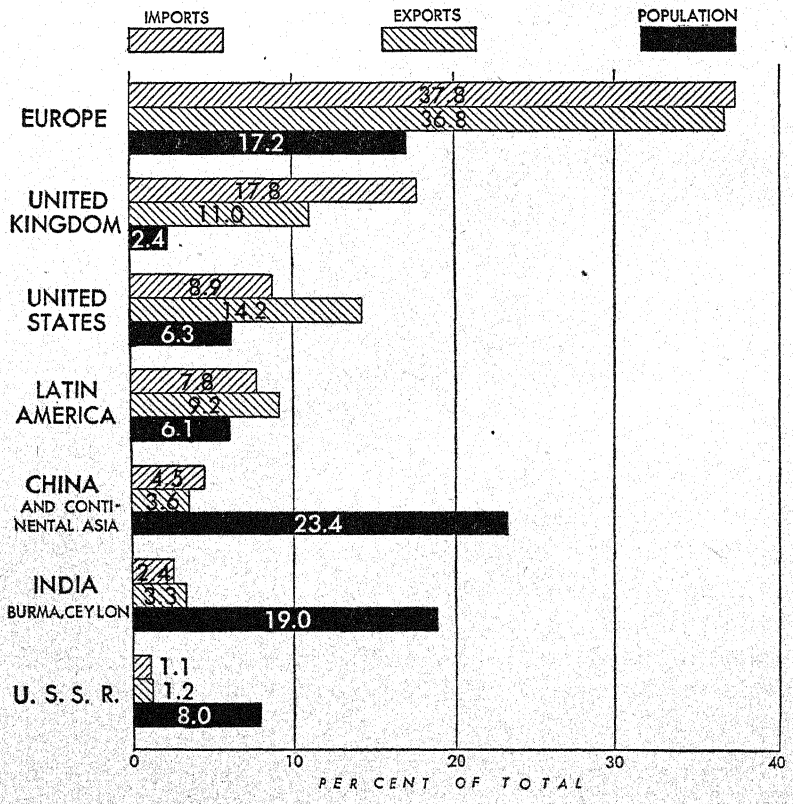
Trade among nations is more than the exchange of commodities and services; it is a network of enormously intricate design which when damaged inflicts great loss to the nations it links together. World War I disrupted world trade but it was virtually re-established by the end of the decade following. The Great Depression altered and disturbed this network to a considerable extent and World War II shattered it. Despite political, economic, and financial difficulties, the network is in process of rebuilding. Patterns of trade normally are relatively slow to change, for trade is governed by the structures of the major economies of the world. The economies linked together by trade cannot alter their individual structures quickly even under conditions of war or depression.

The structures of the industrial economies of Europe, particularly Great Britain, are strongly oriented toward world trade, since they depend heavily upon imports of food and raw materials and exports of finished goods. Great Britain and continental Europe normally account for over half of total world imports and nearly half of total world exports. India, Argentina, Brazil, Australia, and southeast Asia are also heavily dependent upon world trade—as great raw material and food exporters and as importers of finished products. The United States normally assumes a position of importance in world trade next to continental Europe and Great Britain. Before World War II, United States imports were only half those of the United Kingdom, but exports were somewhat larger. In important respects the productive structure of the United States is shaped by this trade.

In 1938 world exports of merchandise amounted to 22 billion dollars. Shipping costs added another 2.5 billion dollars, bringing the total value of imports to 24.5 billion dollars. This was equal to about one third the value of the national output of the United States and equal to the whole of the output of Great Britain or Germany. This trade radiated from the major industrial areas; about 80 per cent of all goods bought and sold entered into the international transactions of Europe, Great Britain, and the United States. These areas embrace only one fourth of the world's population but account for four fifths of the world's trade. In Chart 22 the percentages of world imports, world exports, and world population accounted for by a number of areas are shown.

The composition of the world's merchandise trade is usually classified as (1) foodstuffs, (2) raw and semimanufactured materials, and (3) manufac-

Chart 22 **PERCENTAGE OF WORLD IMPORTS, EXPORTS, AND POPULATION, 1938**



SOURCE: NETWORK OF WORLD TRADE, LEAGUE OF NATIONS, GENEVA 1942

tured goods. Over one fifth of international trade in 1938 was in foodstuffs, much of which originated in Latin America, Canada, and Australia for shipment to Great Britain and Europe. Nearly two fifths comprised raw and semimanufactured materials, most of which was sold to the industrial nations. Finished manufactured goods accounted for the remaining trade—nearly two fifths—between industrial nations themselves and between them and food- and raw-material-producing regions.

In the network of world trade there are well-established patterns of economic interdependence. The coal and steel of Germany are essential to the economic well-being of the rest of Europe, just as the food and fiber of the central and eastern European states are necessary for German well-being. The postwar disruption of this trade, caused in large measure by the disjointed political administration of Germany by the victors, impoverished not only Germany but many of its neighbors as well. The economic interdependence of the European nations was demonstrated in a difficult and costly way by the vengeance shown Germany. Canadian wheat, Argentine beef, Danish bacon, American and Egyptian cotton, Venezuelan oil, and Australian wool are vital to Britain. British manufactured goods in turn make up a sizeable share of the imports of these countries. United States' dependence upon essential raw material imports was dramatized by the war, and foreign reliance upon American manufactured goods was demonstrated (and exaggerated) by the enormous exports of 1946-47.

The United States in World Trade

Before World War II the United States accounted for about 15 per cent of world exports and 10 per cent of world imports. Its growth to a position of importance in world trade, as shown in Table 70, has been rapid. In 1800 its trade was negligible; it imported 91 million and exported 71 million dollars' worth of goods. By 1900 the United States imported nearly 1 billion and exported more than 1 billion dollars' worth of goods. Its peacetime trade grew substantially until World War I, when the demands of Britain and France for goods caused exports to soar. During the prosperous 1920's it attained a position which, except for the depressed period in the early 1930's, it maintained until the outbreak of World War II. War established the United States a second time as the chief exporting nation of the world, a position which it maintained through 1946 when it shipped abroad some 12 billion dollars' worth of merchandise. Thus its industrial expansion and its relative insularity from the destruction of war have combined to put the United States into the dominating position in world trade.

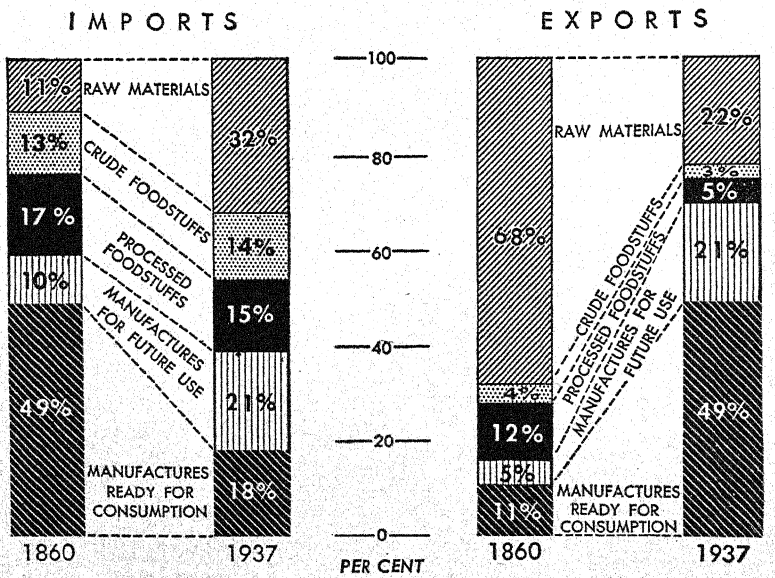
Table 70 FOREIGN TRADE OF THE UNITED STATES

Selected years 1800–1940

Year	Exports of commodities	Imports of commodities	Excess of exports (+) or imports (—)
1800	$ 71 millions	$ 91 millions	$— 20 millions
1820	70	74	— 4
1840	124	98	+ 26
1860	334	354	— 20
1880	836	668	+ 168
1900	1,395	850	+ 545
1920	8,228	5,279	+2,949
1930	3,843	3,061	+ 782
1940	4,021	2,625	+1,396

The composition of America's trade has changed greatly. This can be seen by comparing its imports and exports of the mid-nineteenth century with those of the period preceding World War II, as is done in Chart 23. Foreign trade figures for 1946 are so distorted by wartime exigencies and postwar needs that they are of little value in this comparison. The salient points may

Chart 23 COMPOSITION OF UNITED STATES FOREIGN TRADE, 1860 AND 1937



SOURCE: U. S. DEPARTMENT OF COMMERCE

be summarized from the import and export data for 1860 and 1937. In 1860 nearly three fourths of American exports were raw materials and unprocessed foodstuffs—only one fourth wholly or partly manufactured products. These proportions were reversed in 1937. For imports, 49 per cent in 1860 were finished manufactured goods, and only 18 per cent in 1937; raw materials were only 11 per cent in 1860, but 32 per cent in 1937.

Thus, in mid-nineteenth century the United States was chiefly a raw-material and food exporter, and had become by the time of World War II primarily a manufactured-goods exporter. The expansion of manufacturing in turn altered the character of imports which now are largely (82 per cent) in raw materials, semifinished goods, and foodstuffs. Both trends are likely to be accentuated in the future. There are several reasons for this expectation: (1) American industrial capacity has increased enormously, far more than that of any other industrial and trading nation, and will therefore supply an increased share of world industrial exports; (2) industrial activity consumes large amounts of raw materials; this will tend to reduce (relatively and possibly absolutely) some American raw-material exports; (3) some natural resources in this country are becoming deficient or threaten to become so. Petroleum is a resource which will require importation increasingly in the future. Before World War II the United States was a large exporter of petroleum, but during the war the rapid exploitation of oil deposits to meet military and industrial needs reduced the supply so greatly that the country will probably now become a net importer of petroleum. The expansion of industry coupled with diminishing returns in raw-material production leave no other course open but increased dependence on foreign raw materials.

Besides volume and composition of trade there is the consideration of the direction of American foreign trade, i.e., the relative importance of various markets for exports and sources of imports. Table 71 gives the percentages of United States exports going to the major trade regions in the prewar years of World Wars I and II.

As a market, Europe has declined in importance but nevertheless remains the principal buyer of American goods. Asia, Africa, and South America have grown markedly, from 13 per cent before World War I to 30 per cent before the recent war. The leading exports—comprising two thirds of total exports—going to foreign markets in 1936 are given in Table 72 in their order of magnitude.

The importance of cotton and tobacco should be noted. Not only are they large items in the list, but as exports they comprise a large part of United

Table 71 UNITED STATES EXPORTS, BY PRINCIPAL REGIONS

Regions	*Pre-World War I*	*Pre-World War II*
Europe (including the United Kingdom)	62%	43%
Northern North America (chiefly Canada)	15	16
Southern North America (Caribbean area)	8	9
South America	6	8
Asia	6	16
Oceania	2	3
Africa	1	5
	100	100

Source: U.S. Department of Commerce.

Table 72 UNITED STATES EXPORTS, PRINCIPAL COMMODITIES

Cotton, raw	$361.0 millions
Machinery	334.9
Automobiles	240.2
Petroleum and its products	179.0
Tobacco, unmanufactured	137.3
Chemicals	116.9
Iron and steel products	111.9

Source: U.S. Department of Commerce.

Table 73 UNITED STATES IMPORTS, BY PRINCIPAL REGIONS

Regions	*Pre-World War I*	*Pre-World War II*
Europe (including the United Kingdom)	50%	30%
Northern North America	7	16
Southern North America	14	9
South America	12	12
Asia	15	29
Oceania	1	2
Africa	1	2
	100	100

Source: U.S. Department of Commerce.

States production—33 per cent of tobacco and 56 per cent of cotton production were exported in 1936.

The sources of American imports for the two prewar periods are shown in Table 73.

It will be noted that between the two periods Europe declined in importance as a supplier of goods to this country. Imports from Europe, however, remain largely finished manufactured or semifinished goods. On the other hand, imports from Asia doubled in importance over this period, as did those from Canada. The leading imports before World War II are shown in Table 74.

Table 74 UNITED STATES IMPORTS, PRINCIPAL COMMODITIES

Crude rubber	$158.7 millions
Cane sugar	157.9
Coffee	134.0
Paper and paper manufactures	110.1
Raw silk	102.4

Source: U.S. Department of Commerce.

The United States exports considerably more merchandise to Europe and Great Britain than it imports from this region. An excess of exports is known traditionally as a "favorable balance of trade." In 1936 American exports were 1 billion dollars and imports 700 million dollars in trade with Europe. During the 1920's the figures averaged 2.2 billion dollars for exports and 1.1 billion dollars for imports. An excess of imports over exports is an "unfavorable balance." In trading with Asia, the United States normally has an unfavorable balance; in 1936, imports were 708 million dollars and exports 399 million dollars. Similarly, it usually buys more from the Caribbean area and South America than it sells to them. Asia and Latin America are principal sources of raw materials and foodstuffs and, although they buy manufactured goods in large quantities from the United States, their principal source of supply is Great Britain and Europe—who in turn are the principal buyers of America's manufactured goods exports.

In considering America's import and export positions vis-à-vis these great trading regions, the exchange of *services* as well as merchandise must be calculated. Many services enter into international transactions, mainly transportation services, the services of foreign banks in handling financial arrangements, and the services of marine and other insurers. When British, Norwegian, and Greek vessels carry goods for American account and when London firms insure these cargoes, these services must be purchased by

American businessmen; in 1940 this item cost 327 million dollars. The importance of services will be shown later.

Multilateral versus Bilateral Trade

The *multilateral* character of American foreign trade—and of world trade in general—was touched on previously. American imports from Asiatic countries have exceeded American exports to these countries; on the other hand, United States exports to Great Britain are greater than British imports into this country, and Britain in turn exports more to Asia than it imports from Asia. These relationships are termed *multilateral.* The United States buys jute from India, sells raw cotton to Britain, and Britain in turn sells textiles to India. The United States sells tobacco to the British, who sell cigarettes to the Brazilians, who sell coffee to the United States. These three- (or more) way transactions are a common feature of world trade. When trade is multilateral, countries import from and export to the most favorable markets without limiting purchases from one country to the sales made to that country.

On the other hand, trade may be *bilateral.* If Sweden and the Soviet Union agree to the exchange of given amounts and types of goods over a period of years, the trade between them is regulated and balanced by this governmental arrangement. Or the governments of Argentina and Great Britain may agree that all British purchases in Argentina must be matched by Argentine purchases of British goods. Bilateral arrangements came into prominence during the 1930's because of the financial disturbances of the period; in time, they became powerful instruments of international politics. They represent the growth of political collectivism in the field of international trade and finance, for such arrangements are often made for governmental purposes. The distinction between multilateral and bilateral trade is great in actual effect. Multilateral trade tends to encourage the specialization of productive effort; bilateral tends to discourage it. Under the former, each country can specialize its productive effort to the extent warranted by all markets for its products. The largest export market for coffee is the United States. If Brazilian exports to this country were limited by bilateral agreement to the value of its imports from this country, Brazil would export less coffee to this country. Under a similar bilateral arrangement Brazil would have to curtail its present imports from Britain to the size of its exports to Britain; since the English prefer tea to coffee, it would be difficult to increase exports in this good. Of course, Brazil might shift its purchases of manufactured goods from

Britain to the United States, thus balancing its imports and exports with both countries and at the same time maintaining its coffee sales to this country. This is conceivable, but the arrangement is awkward and inflexible, for bilateralism among nations, as among persons (barter), discourages specialization and the most economic use of resources.

Interdependence of Imports and Exports

Foreign trade is a mutual exchange of commodities and services, but it normally assumes the form of an exchange for money. The two-way relationship between importing and exporting is not always considered when the wage earner or the businessman or the politician, or even sometimes the economist, discourses on the subject of exports and imports. There is a widespread belief that exports are beneficial because the production and export of goods creates jobs, wages, and profits. This is true, but its corollary is not necessarily that imports should be kept as low as possible. The favorable-balance-of-trade idea has enjoyed popular and political acceptance since the days of mercantilism. Until recent decades the advantages derived from a favorable balance of trade were regarded largely in a monetary sense, or more specifically, *in terms of gold flowing into a country.* If a country's exports were valued at, say, 100 and its imports at 75 in a year, then the difference in sales over purchases—25—would be paid in gold. The acquisition of gold—the standard measure of monetary value—was of great importance to countries with no gold mines; and the favorable balance of trade was the principal means of its acquisition. The maintenance of gold receipts is still important to countries with inadequate gold stocks, but it is by no means the overriding consideration that it was in former times.

In the United States, popular opinion continues to stress the desirability of an excess of commodity exports over commodity imports. Such an excess (when adjusted for other transaction between nations) roughly measures the stimulative influence on domestic employment and production. Thus, if in its foreign trade in any given year this country's exports exceed its imports by 1 billion dollars, then this amount constitutes a stimulus to business. So great has been the concern over providing the stimuli to maintain relatively full employment [1] that many economists (and others) have gone beyond the most avid mercantilists of the seventeenth century in their espousal of a favorable balance of trade. Such a stimulus

[1] See Chap. 7 for a summary of this point.

to domestic employment and income may also be a means of encouraging the capital development of other countries.

In other countries, exports are regarded as a means of obtaining *imports*. In England, Sweden, Denmark, and other countries heavily dependent upon imports, the importance of exports rests precisely on this point. When the English export to this country they acquire "dollar exchange"; i.e., American importers must pay for the English goods, and in the financial transaction the English acquire dollars. With these dollars the English are enabled to buy American cotton, machine tools, and other things they want. If they do not want American goods they can use the dollars to buy Canadian or Argentine goods, for these countries can use dollars to pay for purchases of goods from the United States. A country may obtain imports from other countries in the following ways: (1) by the exportation of commodities, (2) by the exportation of gold, which is also a commodity, (3) by rendering services to another country, (4) by borrowing from another country, or (5) by securing gifts (e.g., lend-lease). The principal method by which imports are obtained, of course, is the exportation of goods with the attendant acquisition of money, or an "exchange balance." Great Britain must import to live; but to import she must export commodities and render services (e.g., shipping services), if she is to acquire the money to purchase goods for import. The United States is less dependent upon imports of foreign goods; the varied resources and the great expanse of the country enable it to satisfy most wants by its own productive effort. As a consequence, exports tend to be regarded by most Americans less as a means of buying foreign goods than as a method of "getting rid" of surpluses or of developing new outlets for investment.

But an excess of exports by this country leads to the creation of foreign debts owed to Americans. If these are to be repaid the foreign debtors must acquire dollars in the course of time. Their principal means of doing so is to export more to the United States in the future than they import from the United States. Therefore, America's present export surpluses must be redressed in the future if the debts arising from these surpluses are to be repaid. This shows again that foreign trade is a two-way arrangement; it arises out of specialization of productive activity and the mutual needs of nations. Exports of the United States satisfy foreign wants, and foreign exports to the United States satisfy certain wants in this country more economically than is possible by domestic production.

CHAPTER 35

International Finance

The international transactions of any country fall into two categories: (1) those that require payments *from* other countries, and (2) those that require payments *to* other countries. Exports from the United States to foreign countries, for example, establish a *claim* for payment against the foreign importers, while imports into this country establish a claim against the American importers for payment abroad. Although imports and exports of merchandise bulk large in these claims for payment and obligations to pay, there are many other transactions involving these payments. Services performed between countries, as well as purchases and sales of corporate and governmental securities and of gold and silver, enter into international transactions in the same manner as other commodities. These and other transactions all involve the payment or receipt of money and all enter into a nation's "balance of payments."

THE BALANCE OF PAYMENTS

Those international transactions of the United States for which foreigners must make payment appear as *credit* items in its international balance, and those for which payments must be made to foreigners are *debit* items. Credit items require payment to United States nationals in the form of dollars; debit items, on the other hand, require the payment of foreign money by Americans to foreign creditors. Thus both Americans and the nationals of other countries must normally acquire foreign funds or "foreign exchange" to make these payments.[1]

[1] The American importer may pay the foreign exporter in United States money in many transactions; however, the foreigner's government or central bank acquires this United States money and pays him in his national currency. This procedure does not alter the principle discussed. Later the steps in making payments by this device will be described.

There follows a review of the principal items on the *credit side of the international balance* of the United States in 1938.

1. Merchandise exports amounted to 3.2 billion dollars. This is a credit item because it represents sums owed to American nationals by foreign buyers. These foreign buyers were required to obtain 3.2 billion dollars through international banking arrangements to pay American exporters.
2. Sale of services to foreigners amounted to 1 billion dollars. These services include transportation by American railroad and shipping companies handling the goods. When these shipments are insured by American companies, a service is also sold to foreigners. The total also includes the *services* of American capital invested abroad for which foreigners pay interest and dividends. Services rendered by American banks to foreigners are also included in this total.
3. Sale of long-term corporation bonds, United States government bonds, and corporation stocks to foreign investors amounted to 1.6 billion dollars. In the international balance this type of transaction is a sale in all respects similar to the sale of goods and services. This transaction is in a sense an export of United States securities, mostly corporate.
4. Sale of short-term securities to foreigners amounted to nearly 1.2 billion dollars. This type of sale is similar in all respects to the sale of long-term securities.
5. Foreign bank holdings of dollars (their bank deposits in American banks) increased by 117 million dollars during this year. This appears as a credit item because foreigners acquired dollars in this amount, just as they must to pay for the goods, services, and securities they purchase from American nationals.
6. All other credit transactions amounted to 1 billion dollars.

Credit items totaled 8.2 billion dollars, an amount which had to be paid in American dollars. Approximately 40 per cent of these payments arose from the foreign purchase of American goods, and over 30 per cent from the purchase of American securities.

On the *debit side of the international balance* of payments of the United States, the following items may be summarized:

1. Imports of foreign goods into the United States amounted to 2.3 billion dollars. Since the foreign sellers finally require payment in their

own national currencies, American buyers or their agents had to convert 2.3 billion dollars into foreign money to make payment.

2. Purchase of foreign services amounted to nearly 1.3 billion dollars. This total included shipping services, insurance services, the expenditures of American tourists abroad, and payments of interest and dividends by American corporations for the services of foreign capital invested in this country.
3. Purchase of foreign securities by American investors amounted to 1.5 billion dollars. This in a sense is an import of foreign securities.

These three items add up to 5.1 billion dollars, an amount of American money which had to be converted into pounds, francs, marks, kroner, and other currencies in order to make payment.

Consider now the totals in the transactions described thus far. The list of credit items showed foreigners owing persons or businesses in the United States a total of 8.2 billion dollars. The list of debit items showed Americans owing foreigners 5.1 billion dollars. This is a difference of 3.1 billion dollars which was balanced by gold and silver shipped from abroad to the United States. Thus, to complete the list of principal *debit* items there must be added:

4. Imports or purchases of gold and silver, amounting to 3.1 billion dollars.

Thus credit and debit items both amounted to 8.2 billion dollars, and the accounts balanced.

The position of gold and silver (particularly gold) has special importance. Had American nationals bought 3.1 billion dollars additional commodities, services, and securities in 1938, the balance would have been struck. But presuming that Americans had bought all they wanted of these items, or that the foreigners sold all of these items they were able, to balance the international account it was necessary for the foreigners to sell something else this country was prepared to buy. That something else was gold (and silver in smaller amounts) because the United States stands ready to buy all gold offered to it at the price established by its own federal law. This transaction can also be expressed in financial terms. Foreigners had to acquire more than 8 billion American dollars to pay their debts, for exporters in the United States do not normally accept foreign money. Thus sales by this country created a *foreign* demand for American dollars to make payment. In what way do foreigners obtain this money? This is evident from a glance at the other side of the balance. The United States bought com-

modities, services, and securities amounting to 5.1 billion dollars; this money was surrendered to foreigners to obtain *their* money to pay American debts. Thus American purchases provide foreigners with American dollars (or dollar exchange or dollar balances) with which they paid part of their debts. For the remainder they shipped gold and silver which this country purchased for 3.1 billion American dollars. The total amount of dollars thus transferred to foreigners by American purchases of commodities, services, securities, and gold amounted to 8.2 billion dollars. With this they paid for the things they bought from this country, and added 117 million dollars to their bank balances here.

International transactions thus involve an exchange of commodities (including gold and silver), services, and securities (evidence of debt and of ownership), as well as an exchange of different kinds of money. Exports from this country require foreigners to obtain dollars, which they do by exchanging their pounds, francs, kroner, and other monies for dollars. American imports require that importers obtain these kinds of money, which they do by giving dollars to foreigners in exchange. In short, each party to the transaction needs the other's money, just as each party wants the other's commodities, services, or securities. Since every purchase by one party is a sale by the other party to the exchange, international transactions or payments necessarily balance. The transfer of money in this process is effected by banking houses engaged in international financial operations.

It has been said that in 1938 the United States exported 3.2 billion dollars of merchandise and imported 2.3 billion dollars. This may be termed a *favorable balance* in its *commodity* trade with the rest of the world. In its trade with individual countries, however, the United States had some unfavorable balances (for example with Asia), which were more than offset by favorable balances with others (such as Great Britain). These regional differences represent the varying degrees and types of interdependence between this economy and the other parts of the world. As was shown in the table of exports and imports for the period from 1800 to 1940,[2] a favorable commodity balance has been a feature of the American trade position for many decades. This has been partly offset by the purchase of more services abroad than have been rendered to other nations by the United States. These include services rendered to American tourists, who spent large sums of money abroad in the two decades between the world wars.[3]

[2] See Chap. 34, p. 746, Table 70.

[3] Tourist expenditures are debit items and involve the purchase of services and goods abroad. Instead of purchasing foreign goods as imports into this country, tourists go abroad to consume them.

Nevertheless the United States has enjoyed a relatively "strong" or "active" balance of payment position throughout most of these years, as evidenced by gold imports into this country and by extensive purchases of foreign securities by American nationals.

It is instructive to compare the balance of payments in 1938 with that in 1946. The 1946 balance of payments of the United States given in Table 75 is in somewhat more detail than was given for 1938. Also, it will be noted that aggregate transactions in 1946 were twice those of the earlier year, partly because of price increases but also because of greater shipments of goods.

Table 75 UNITED STATES BALANCE OF PAYMENTS IN 1946

(in billions of dollars)

A. *Credit Items or Receipts from Abroad*		
1. Merchandise exports	12.1	
2. Sale of services to foreigners	2.6	
3. Services of United States capital (interest and dividends)	.6	
4. Sale of United States owned foreign capital	1.0	
5. Unilateral transfers (grants and gifts)	.2	
Total receipts		**16.5**
B. *Debit Items or Payments*		
6. Merchandise imports	5.3	
7. Purchase of foreign services	1.7	
8. Interest and dividend payments (capital services)	.2	
9. Purchase of foreign securities	4.0	
10. Purchase of foreign-owned United States securities	.3	
11. Unilateral transfers (grants and gifts)	3.3	
Total		**14.8**
12. Purchase of gold	.6	
13. Increase in United States bank balances abroad	.3	
14. Decrease in foreign bank balances in United States	.8	
Total		**1.7**
Total payments		**16.5**

Source: Survey of Current Business, March, 1947.

Merchandise exports and imports (items 1 and 6) need no further explanation. Services sold and purchased (items 2 and 7) are comprised principally by transportation and the costs of travel. Interest and dividend receipts and payments (items 3 and 8) are shown separately for 1946 but were included under "services" in the analysis of the 1938 balance.

Capital items in the tabulation above need some explanation. In the list of credit items is shown the sale of 1 billion dollars of American-owned capital and securities (item 4) which had, until that time, represented United States investments in foreign countries. That is, American investors with holdings in foreign countries sold many of these holdings back to foreigners. The Canadian and Argentine governments, for example, redeemed large bond issues which had been purchased by Americans in earlier years. The Argentine government purchased its American-owned telephone system. Other transactions of a similar nature lead to the payment of dollars by foreigners to United States nationals.

In the debit items on long-term investment, there is an extensive purchase of foreign securities—4 billion dollars (item 9) by this country. The bulk of these investments, amounting to 3.3 billion dollars, were made by the United States government, including Lend-Lease credits, Export-Import Bank credits, loans to foreign governments for the purchase of United States surplus property abroad, the payment of a large sum on the 3.5 billion-dollar United States loan to Great Britain, and other loans and advances by the government. Private purchase of foreign securities in 1946 was less than the sale by Americans of their capital holdings abroad. It will be noted also that American nationals purchased American securities held by foreigners (item 10). These capital transactions of a debit nature involved the payment of dollars by the government—and to a lesser extent by private investors—to foreigners; these may be described as the importation of foreign securities and notes.

The 1946 balance statement introduces a new item: unilateral transfers, which are the transfer of funds as mainly gifts and contributions. In 1946 this country contributed 3.3 billion dollars (item 11), mostly by the government, for relief and rehabilitation in many parts of the world. On the credit side of the balance appears 200 million dollars (item 5) which represent in part repayment or reimbursement of such grants by foreign countries. Thus the net amount granted as gifts—or other transactions of dubious repayment prospects—aggregated 3.1 billion dollars on the part of the United States in 1946. This payment to foreigners financed about 20 per cent of their purchases of American commodities and services (appearing among the credit items).

The debit items 6 through 11 provided foreigners with 14.8 billion dollars. Their purchases from this country (the credit items) amounted to 16.5 billion dollars. The balancing of accounts was effected by items 12, 13 and 14. Gold imports or purchases, item 12, have been described already. Item 13 is the increase in American bank balances in foreign countries. To

acquire additional balances of pounds, francs, and other foreign money, American banks exchanged 300 million dollars for these funds, thus providing foreign banks and governments with this amount of American money. Foreign banks have large bank deposits in this country. These were reduced by 800 million dollars (item 14) in 1946. Debits and credits therefore balanced at 16.5 billion dollars.

In Table 75 it will be noted that in 1946 the United States sold commodities and services valued at over 15.3 billion dollars; this covers the first three of the credit items. Foreigners acquiring these commodities and services had to obtain dollars to pay for these things. The debit items indicate the ways in which they obtained these funds. Table 76 transposes the debit items of Table 75.

Table 76 SOURCES OF FUNDS

Method of obtaining dollars	*Billions of dollars*	*Per Cent of total*
1. Sale of goods and services to the United States (items 6, 7, 8)	7.1	46
2. Liquidation or sale of foreign-owned capital assets and gold to the United States* (items 10, 12, 13, 14)	2.2	14
3. Foreign borrowing from United States (item 9 minus item 4)	3.0	20
4. United States unilateral transfers in money and commodities (item 11 minus item 5)	3.1	20
Total	**15.4**	**100**

Source: Survey of Current Business, March, 1947.

* Item 2 above includes gold (600 million dollars), decrease in bank balances (900 million dollars), increase in United States bank balances (300 million dollars) and purchase (or redemption) of United States securities (300 million dollars). Slight discrepancies in figures are due to rounding.

INTERNATIONAL TRANSACTIONS

The flow of money payments between countries is handled through large banks in the financial centers of the various countries, such as the Chase National Bank, the National City Bank, and the Guaranty Trust Company in New York, and by other private banks in London, Amsterdam, Stockholm, and elsewhere. Financial transactions between governments are carried out by the central banks such as the Bank of England, the Bank of France, and the Federal reserve banks (especially the New York Federal

Reserve Bank). These banks form an international financial network which links together all the currencies of the world. In this country, banks in all parts are linked to the large New York banks through correspondent relationships;[4] and in each country abroad other banking systems tie virtually all banks to those particular banks engaged in foreign finance. Thus a person in St. George, Utah, can send funds to, or receive payment from, a person in St. Gall, Switzerland, through local banks which have a correspondent relationship with large banks.

The large banks which deal in international finance maintain deposits in foreign banks. The New York banks hold deposits of English, Swiss, French, Swedish, Dutch, and other banks from all parts of the world; and foreign governments also have bank deposits in New York.[5] Foreign bank and foreign government deposits in banks of the United States amounted to more than 6.5 billion dollars in 1946. These are chiefly demand deposits against which checks may be drawn when foreigners make payments for American goods, services, and securities. Deposits to the credit of foreigners are termed their *dollar balances* or *dollar exchange.* Foreign deposits in the United States have grown rapidly. In the mid-1930's they stood at less than 1.5 billion dollars; in 1946 they were more than 6.5 billion dollars, the increase occurring largely during the war years. The presence of such large foreign holdings of dollar deposits reveals the dominant financial and economic position of the United States.[6]

American banks in turn have deposits in English, Swedish, French, Latin-American, and other banks. These deposits are in the currency of these countries; thus American banks have deposits in pounds sterling in London, in kroner in Sweden, etc.; they are referred to as the "sterling balance" (sterling exchange), "kroner balance" (kroner exchange), etc. Late in 1946 the deposits of American banks in foreign countries had a value in foreign money equivalent to 600 million dollars at the prevailing rates of exchange.

These international checking accounts or demand deposits play a dominant role in the conduct of international economic and financial transactions. Checks (or drafts) are constantly being drawn against them to make

[4] Described in Chap. 10.

[5] This discussion mentions only foreign deposits held in the New York banks and New York bank deposits in other countries. This is for convenience only. Some large banks in cities other than New York also hold foreign deposits and have their own deposits abroad. New York, however, is the principal center of international finance.

[6] The size of some of the deposits may be mentioned. Canadian deposits were 1.1 billion dollars. Latin-American deposits 1.1 billion dollars, British 625 million dollars, Swiss 350 million dollars, the Netherlands 225 million dollars, French 250 million dollars, Asia 1.4 billion dollars. These deposits of course change as payments are made and receipts accepted.

international payments, and checks or drafts are constantly being deposited in them as payments are received. In essential respects these international banking arrangements are similar to domestic banking arrangements in that they permit payments and deposits to be made by means of checks or drafts. In the international sphere these checks are called drafts, or foreign exchange. In principle they differ in no basic respect from checks drawn by one local businessman in payment of a debt owed to another local businessman.

Payments Received from Foreigners

Suppose an American manufacturer of oil burners receives an order from a Swiss firm in St. Gall amounting to $50,000. This American export transaction may be financed in several ways, but probably the financial arrangement would be as follows:

The Swiss importer arranges with his bank in St. Gall to obtain a letter of credit in New York amounting to $50,000; on his part he agrees to pay to his bank an amount of Swiss francs equal to $50,000. The St. Gall bank has a deposit of its own in the Guaranty Trust Company of New York, or if not it would arrange with a large Swiss bank that did have a deposit in New York. The letter of credit assures the American manufacturer that he will be paid $50,000 by checks drawn on the Guaranty Trust Company when he delivers the oil burners for export to the Swiss purchaser. The American manufacturer after delivering the burners for shipment thereupon draws a bill or draft against the Swiss importing firm that he presents to his own bank which, it may be supposed, is also the Guaranty Trust Company. The letter of credit authorizes the Guaranty Trust Company, or an agent in New York handling such matters for the Swiss bank (or the Swiss importer), to draw a check for $50,000 payable to the manufacturer at the time delivery of the burners is made. This is done and the manufacturer deposits the check; the Guaranty Trust Company credits his deposit account by the amount of the check and debits the deposit of the Swiss bank. The manufacturer has received his money and has completed the transaction. Meanwhile, the burners are shipped to Switzerland, either by the manufacturer himself or an American export firm or an agent of the Swiss importer. If the burners are placed on board an American ship and insured by a New York marine insurer, payment must be made by the Swiss importer for these services by means of a check drawn against the Swiss bank's deposit in the Guaranty Trust Company.

There are several possible variations to this procedure, but the eventual

result in the financial transaction is the same. The export of United States goods (oil burners) and the rendering of services (transportation and insurance) by American nationals involved a *reduction* (by $50,000) in the foreign dollar balances in New York. The $50,000 bank deposit is thus transferred from a foreigner to an American (the exporter).

Other illustrations of international transactions may be given. Suppose a Londoner wants to buy some General Motors stock on the New York Stock Exchange. He arranges with the Midland Bank of London to pay for these securities in New York by a check drawn on its deposit with the Chase National Bank. Midland authorizes the Chase National to draw a check against the Midland deposit to cover the purchase of the shares. With this authorization assured, an agent of the British bank or of the London purchaser buys the stock through a broker on the Stock Exchange. The seller of the stock (in effect the broker) deposits the check in his own bank; when the check is presented for collection to the Chase National it debits the deposit of Midland Bank. This transaction, too, can be handled in other ways, but the method described once more shows how the sale (or export) of American securities reduces the bank deposits held in New York by foreign banks.

Another method of financing exports should be described. Suppose an American exporter of cotton ships to a British importer a consignment worth £10,000. The exporter (drawer) draws a bill for this amount against the importer (drawee), directing him to pay this amount to a designated third party (payee) in London. The third party is the National City Bank, which has a deposit in the Midland Bank of London. The exporter takes the bill to his bank in New York, the National City Bank, together with the bill of lading and papers covering the insurance of the cotton. (A bill of lading is a document issued by a transportation company receipting for goods for shipment to a designated point.) The National City Bank buys the bill drawn against the English firm, paying the exporter $40,000. The exporter now has his money, and his bank has the bill and shipping documents (bill of lading and insurance papers). These are sent to the Midland Bank in London while the cotton is in transit. When Midland receives the bill and documents it presents the bill to the importer for collection. The importer pays the Midland Bank £10,000, receives the bill of lading properly endorsed, and obtains his cotton. The Midland Bank thereupon credits the deposit of the National City Bank for £10,000.

In this type of transaction the export of goods from this country created an additional sterling balance for the New York bank in London. At one time a large part of American exports was financed in the manner just

described. In recent times the method which reduced dollar balances of foreign banks (as in the export of oil burners) has become prevalent. The latter is simpler and safer for the American exporter because the buyer has to take the responsibility of making payment *before* the goods are shipped. There are two important causes of this change in the method of financing United States exports: (1) the United States has become the principal industrial producer in the world; thus, its exporters can now largely state the financial terms, and (2) the large dollar balances held by foreigners in the United States are a convenient means of payment for American exports.

The financing of American sales to foreign buyers may now be stated in summary form.

1. If these transactions follow the procedures of the Swiss illustration, they involve:
   a. *A reduction of foreign bank deposits (in dollars) held in American banks;* and
   b. *An increase in reserves of foreign banks in their own monies;* i.e., the foreign bank gives up dollars in exchange for additions to its reserves (or reductions in its deposit liabilities).
2. If the transactions conform to the cotton export procedure, the American bank:
   a. *Gives up dollars (adds to its deposit liabilities) to the American exporter;* and
   b. *Acquires foreign money in deposits abroad,* i.e., it exchanges dollars for foreign money.

Thus payments received for American sales abroad have the effect either to (1) increase American banks' holdings of foreign money, or (2) decrease foreign bank holdings of American money.

Payments Made to Foreigners

Debit transactions reverse the flow of payments—from American nationals to foreigners. The methods of payment, however, show considerable differences.

Suppose an American importer orders $10,000 worth of woolens from an English firm. If both parties to the transaction have been doing business with each other for many years the English exporter may simply ship the goods and draw a bill against the American importer for payment of £2,500. To obtain possession, the American importer buys a sterling draft for this amount from his bank, the National City Bank. He pays $10,000

(plus a service commission) to the bank and receives its draft for £2,500, drawn on its deposit with the Midland Bank and payable to the English importer. This he sends to the English firm, which presents it for collection to the Midland Bank in London, which in turn debits the National City Bank deposit for the amount. The National City Bank in effect converts £2,500 of its London deposit into $10,000 in New York. Its sterling balance declines but its reserves in New York increase; or, if the importer has a deposit in this bank the bank's deposit liabilities are reduced.

Payment may take a different course if the English exporter demands payment in London. He then draws a bill against the American importer; the American firm meanwhile must obtain a letter of credit from the National City Bank enabling the English exporter to obtain payment from the Midland Bank before the goods are shipped. This procedure is precisely the same as followed in the oil-burner export from the United States, except that now it is an English exporter and an American importer. In this case also the National City Bank deposit in London is reduced, while its funds in New York increase (or its deposit liabilities decline) through the payment by the American firm to meet its commitment (i.e., the American importer "buys sterling" from his bank).

In recent years American importers have increasingly paid for their imports in dollars, and not by means of a foreign draft. This procedure is as follows: A Turkish tobacco exporter receives an order from an American buyer. The tobacco is shipped and the Turkish firm draws a bill against the American importer, directing him to pay the amount of dollars involved to the Chase National Bank to the account of a Turkish bank. Before the American importer can take possession of the tobacco he must pay the Chase National. As a result, the importation of goods into this country increases the bank deposits held by the Turkish bank in New York. The Turkish exporter meanwhile is paid by the bank in Turkey in Turkish funds. The bank relinquishes its national money and acquires American money in the transaction. It may be seen that the two procedures which have been summarized for United States exports (pages 761-62) apply, inversely, for United States imports.

Procedures similar to this have been followed widely in recent years, with the result that foreign bank balances have grown enormously in New York. There are a number of reasons for this development. In years past London banks and the banks of the major European countries attracted such balances; wartime conditions made this physically impossible or at least risky. The United States was both safe and convenient for such balances. Moreover, the huge imports of foreign raw materials into this coun-

try during the war, especially from Latin America, and the shortage of American goods for commercial export gave the Latin-American countries the choice of leaving their balances in New York banks, converting them into gold, or buying American securities. The building up of bank balances or dollar exchange in New York was encouraged by the expectation that large purchases would be made in the United States at the end of the war.

Finally, then, the effect of the transactions on bank balances, i.e., on a nation's foreign exchange, may be summarized as follows:

1. American exports of goods, services, and securities either
   a. *Reduce foreign bank deposits held in the United States,* or
   b. *Increase the deposits of American banks held in foreign countries.*
2. American imports of goods, services, or securities either
   a. *Reduce balances of American banks on deposit abroad,* or
   b. *Increase foreign bank deposits held in the United States.*

Some Complicating Factors

The discussion thus far has assumed that trade and finance involve only two countries. In practice, of course, this need not be the case. Consider, for example, the following situation: Colombia exports more to the United States than this country exports to Colombia. As a consequence the deposits of the Colombian banks in New York are larger than are needed for the purchase of American goods. What can the Colombian banks do with this "surplus" of dollars? They can use these balances to pay for the goods which Colombians buy from Great Britain. Colombia buys more from Britain than it sells to Britain and therefore has a "deficit" in its British trade. Colombian importers therefore buy dollar exchange from their banks and pay their debts to the British in dollars. Why do the British accept dollars from Colombian importers and not insist upon sterling? Because the British need dollars to pay for their imports from the United States since these imports normally are greater than exports to the United States. This is termed *triangular exchange.* Because the American dollar has attained world dominance, it is used widely in transactions throughout the world. It has become something of an international currency, just as the pound sterling has been used (and still is to a lesser extent) as an international medium of exchange.

International finance is complicated further by regulations of many governments in their control over financial operations. An illustration will indicate the nature of the controls. In country A a firm wishes to sell

goods to an American importer. In order to get a license from his government to export to the United States, the exporter must agree (1) to require payment from the American importer in United States money, and (2) to require that payment be made by the American buyer to the New York account of one of the banks of country A named by the government. This bank in turn pays the exporter in its national currency. In this way the government of country A builds up dollar deposits for its banks in New York and may also specify what kind of goods may be bought by its firms. Suppose now that this foreign firm wishes to buy goods from the United States. Its previous exports to the United States created the dollar balance which could be used for this purpose. In the absence of further regulations it would be able to buy a dollar draft from the bank to purchase American goods. However, country A may control this type of transaction too, by authorizing its banks to sell dollar exchange only for specified purposes and in specified amounts. In effect it rations or allocates the dollar balance of its banks. Moreover, it is clear that the government can also control the type of goods bought by its firms by controlling sales of dollar drafts. These and other means of control are treated more fully in Chapter 36, but it is important here to indicate that the flow of funds is subject to controls in most countries.

FOREIGN EXCHANGE RATES

International financial transactions are carried on with certain rates of exchange existing between the currencies involved. In one of the foregoing illustrations, an American cotton exporter sold to his bank a bill which he had drawn against the English buyer. The bill was £2,500; the bank paid the cotton exporter approximately $10,000 for it, and the bank's deposit in London was subsequently increased £2,500. Also, when the Swiss importer arranged a letter of credit in New York he paid his bank in St. Gall a certain sum in Swiss francs; in effect he purchased a dollar draft with his francs. The cotton exporter is vitally concerned about the amount in dollars he will receive for the sterling bill, just as the Swiss importer is interested in the number of francs he must pay to obtain dollars in New York. In both cases the amount depends upon the rate of exchange—sterling for dollars, and dollars for francs.

The rate of exchange is the price paid in one country for a unit of foreign money payable in the foreign country. In 1947 the price of the pound sterling in New York was approximately $4.03, with the sterling draft payable in England. The moneys of virtually all foreign countries may be pur-

chased in the New York foreign exchange market at certain prices, or rates of exchange. Thus in 1947 the Swedish kroner cost 28 cents, the Swiss franc 20 cents, the French franc 0.9 cent, etc. The rate of exchange may also be expressed as the number of units of a foreign money that may be purchased with a dollar; for example, $1 buys 44 Belgian francs.

The Gold Standard

Under gold-standard conditions the rates of exchange are related to the amounts of gold contained in the various monetary units. In 1931 the British sovereign contained 113.00116⅔ grains of fine gold, and the dollar contained 23.22 grains of fine gold. The ratio was $\frac{113.00116\frac{2}{3}}{23.22}$ or 4.86+. That is, the British sovereign contained 4.86+ times as much gold as the dollar. This established the *mint par of exchange*. The German reichsmark contained about one fourth as much gold as the dollar; therefore $1 = 4RM$ (or $1RM = 25$ cents) was the approximate mint par of exchange in this case. By the same calculation the British sovereign (sterling) equaled approximately 19.5 German marks. The mint par of exchange is not the rate of exchange. This latter rate is the purchase price of a foreign money. But if the country is on the gold standard, the purchase price normally fluctuates about the mint par, depending upon the supply of the foreign money and the demand for it. When the foreign exchange rate coincides with the mint par the rate is said to be "at par."

Exchange rate determination under gold-standard conditions can be illustrated by a description of British-American experience in the decades before World War I and in the period 1925-31. (In 1931 England abandoned the gold standard and many other nations followed suit.) During these two periods, trade practice required American exporters to draw bills on their English buyers (as in the cotton-export illustration), and American importers bought sterling drafts from their banks to pay for the goods they purchased from England. Thus, exports created a *supply* of sterling exchange in London to the credit of American banks. Imports, on the other hand, meant a *demand* for sterling drafts or exchange and a consequent reduction of New York bank deposits in London. When imports and exports balanced, the amount of exchange purchased by the banks from exporters was equal to the amount sold to importers, and the rate of exchange approximated $4.86, or par.[7] Under these circumstances the

[7] The banks' buying rates were below the selling rates by a small amount. This yielded the profit to banks for the handling of this business.

sterling deposits of the American banks held in the London banks remained stable.

When exports *exceeded* imports, the sterling balances to the credit of American banks increased and the rate of exchange dropped below par; i.e., as the supply of sterling exchange increased its price fell. The sterling balances increased because the amount of sterling offered by exporters and purchased by the banks was greater than the amount demanded by importers. As this occurred, the banks' cash resources in New York were diminished; i.e., the money paid out to exporters for sterling bills exceeded the money the banks acquired in selling drafts to importers. Large balances in London banks were not desirable for the American banks, nor obviously was the diminution of their cash resources. Yet even under these circumstances the New York banks would continue to buy sterling bills offered to them by exporters—but at a price somewhat below par. These purchases were not made to increase their already plentiful London balances, but rather for the purpose of using the additional sterling in London to purchase gold for shipment to New York. This, however, entailed a cost including freight and insurance. The rate paid for sterling, therefore, was par minus the cost of bringing gold from London to New York. This rate was termed the *gold import point.* In London the American banks could buy gold worth $4.86 in New York, but it would cost about 3 cents to bring this amount into the country. Thus, American banks paid $4.83 for sterling bills purchased from exporters and used the London credit thus established to bring gold to this country. For example, a bank would buy a £10,000 bill for $48,300; with this amount of sterling, gold was purchased in London worth $48,600 in New York; shipment, however, cost $300, so the bank broke even on the transaction. By the importation of gold when the supply of sterling was large (because exports and other credit items were large), American bank balances abroad were kept at the desired level and the sterling rate was prevented from falling below the gold import point. Thus the banks bought all sterling offered to them but the price paid did not fall below the gold import point.

On the other hand, when *imports* exceeded exports, sterling balances to the credit of American banks declined and the rate of exchange rose above par. The decline in sterling deposits in London occurred because the amount of sterling demanded by American importers exceeded the amount offered to the banks by American exporters. Adequate balances of course were needed if the banks were to sell the exchange demanded from them by importers. As balances declined banks increased the rate of exchange. When the rate increased to $4.89 the banks found it profitable to buy gold

in New York and ship it (or gold they had in their vaults) to London to replenish their sterling balances in the London banks. Again, the cost of shipment was the deciding factor. Gold worth $4.86 in New York (and £1 in London) cost nearly three cents to ship to London; therefore sterling balances could be replenished for $4.89 per pound. This rate was termed the *gold export point.* Shipment of gold from the United States enabled the banks to sell all the sterling exchange demanded of them at a rate that did not exceed the gold export point.

Under gold-standard conditions, then, the foreign exchange rate varied from the par of exchange by no more than the cost of shipping gold, which was less than 1 per cent above or below par. *The rate of exchange fluctuated within these gold shipping points in accordance with the supply of and the demand for exchange.* It may be seen that the great virtue of the gold standard was this stability in the rate of foreign exchange achieved, in a sense automatically, by gold shipments.

During the periods under discussion, before World War I and in 1925-31, most of the world was on the gold standard. All moneys were linked directly or indirectly to it. The operation of the international gold standard of course was not so simple as described here. A favorable balance of payments for the United States in its trade with Britain did not necessarily mean the shipment of gold to the United States. The surplus sterling balances of American banks often were used, not to acquire British gold, but to pay Brazilian or Malayan exporters of goods to this country, who in turn used sterling drafts to pay their debts to the British. Gold was merely the *final balancing item* in the balance of international payments for all countries linked to the gold standard.

The Abandonment of the Gold Standard

Gold performed the important task of stabilizing exchange rates throughout the world so long as gold shipments did not strip a nation of its needed monetary reserves.[8] But a heavy loss of gold threatened the monetary stability of the nations suffering this loss. This occurred in Britain in the early 1930's. Although the conditions which led to the abandonment of gold were complicated by many factors which cannot be described at this point, it may be said that during the 1930's Britain suffered financial difficulties which weakened the position of the sterling exchange as an international medium. Foreign banks began to decrease their deposits in London banks, making

[8] See Chaps. 8 and 11 for the discussion of the importance of gold in a nation's monetary and banking system.

their withdrawals in gold. Because of the heavy drain of gold from London banks by France, the United States, and other countries, sterling could no longer be supported by the export of gold. Britain abandoned gold, and the sterling rate of exchange fell far below par (in 1932 to $3.50).

Britain abandoned the gold standard because her supply of gold reserves was threatened. But while some countries—and almost all major ones except the United States and France abandoned gold at about this time—also sought to protect their gold reserves, others took this action to *depreciate their currencies* to gain an advantage in foreign trade. The general decline in world trade beginning in 1930 caused great difficulties in countries heavily dependent upon exporting. As they saw their export sales dwindling, some sought to arrest the decline or reverse it by currency depreciation. Depreciation was expected by these countries to increase their exports, in a manner such as follows: Suppose country A's currency had been valued in the New York foreign exchange market at 50 cents. If, by abandoning gold and taking other steps (notably exchange control) to cheapen its currency, country A succeeded in getting the rate down to 35 cents in the New York market, then American importers could buy country A's foreign exchange at a 30 per cent discount. If the prices of goods in country A remained unchanged, American importers obtained these goods at a 30 per cent reduction in American money. As a consequence American imports of its goods might increase substantially—to the benefit of productive activity in country A. It was this expectation that encouraged some exchange depreciation in the early 1930's.

Often, however, this practice defeated its own purpose. If foreign buyers thought the exchange rate would decline by more than 30 per cent they would refrain from buying in the hope of getting a still better bargain later on. If so, country A's exports might continue to drop and its economic situation continue to worsen. Or prices within country A might rise; should they rise by 40 or 45 per cent its goods would cost American importers just as much as they did before depreciation and price inflation, thus offsetting the advantage of the lower rate of exchange. Moreover, if other countries depreciated their currencies by a like amount, country A would be competitively in no better a position than before. If the cheapening of currency becomes a matter of keeping ahead of the next country, the process can lead into a bottomless pit.

These international financial troubles were not only mainly caused but also subsequently intensified by the decline in foreign trade and foreign investments; in turn, they contributed to these declines. World War I debts owed to the United States by several countries and reparation payments

owed by Germany to its victors added further burdens to the international financial system. During the first years of this international financial disorder the United States maintained its traditional gold-standard position. But depression was deep in this country and its banking system was suffering heavily from thousands of bank failures. The financial crisis in the United States led to (1) the hoarding of gold internally, and (2) the export of gold, especially to Great Britain. Together these developments led to the abandonment of gold in this country in 1933.

As already described [9] this was followed by the nationalization of gold stocks and an increase in the price paid by the government for gold, i.e., the depreciation of the dollar in the international market was accomplished by raising the price of gold above the mint price of $20.67 a fine ounce. Early in 1934 the dollar was legally devalued by Act of Congress and the mint price of gold was stabilized at $35 an ounce. The dollar accordingly was devalued 41 per cent; dollar depreciation was at an end and the United States adopted a gold standard so far as concerns the international sphere. Devaluation was undertaken on the belief that prices in general would rise in correspondence with the rise in the price of gold; the price of gold was increased approximately 70 per cent. The principal consequence of devaluation was a great influx of gold, attracted by the high price paid by the United States Treasury.

Abandonment of gold by the United States caused a sharp fall in the price of the dollar in foreign exchange markets abroad and by the same token a rise in the rates of foreign exchange in this country. Sterling, which had dropped to $3.50 in 1932, rose to nearly $6.00 in the New York foreign exchange market. This rise represented a strong demand for sterling balances and something of a "flight" from the dollar. Since gold exports from the United States were prohibited by law during the period March, 1933, to February, 1934, the demand for sterling caused the rate to exceed the old gold export point—just as England's abandonment of the gold standard in 1931 stopped gold coming into the United States when the rate dropped below the gold import point (or England's gold export point).

The breakdown of the gold standard disrupted the system of foreign exchanges and the world-wide structure of foreign exchange rates. The question now arises: In the absence of gold pars and the stabilizing influence of gold movements, what is it that determines rates of exchange between countries? Rates may be left free to "find their own level," or they may be regulated by governmental policy. These are not mutually exclusive alternatives, as shall be shown in the following discussion. But attention will be

[9] See Chap. 11, pp. 256-58.

given first to the influences at work when rates vary without either gold or government control as the stabilizing factor.

Purchasing Power Parity

The rate of exchange between the currencies of two countries on "paper" standards may be explained in terms of "purchasing power parity." An exchange rate that represents or measures purchasing power par is one that expresses an *equivalence in the purchasing power of the currencies in both countries.* Suppose, in the absence of gold and governmental control, the rate of exchange between the pound and the dollar was £1 = $4. If commodity prices in the two countries were in equilibrium, this would mean that one pound could purchase four times as much in Britain as one dollar could in the United States. Or four dollars in the United States would buy about as much here as it would in England when converted into pounds sterling at £1 = $4.

The maintenance of par in this sense is possible only if both countries enjoy relatively stable price levels and are otherwise pursuing relatively tranquil ways. The foreign exchange rate would then tend to find a level which would equate the respective purchasing powers of the two sums thus exchanged. Purchasing power depends upon prices; therefore a change in the general price level in one country would alter the rate of exchange. The £1 = $4 rate is appropriate only to given price levels in both countries. Suppose now that the British price level were to rise 100 per cent. The purchasing power of the pound would be halved if no other intensifying or offsetting change occurred. If the American price level remained unchanged, the foreign exchange rate would tend to settle at £1 = $2 which is the new equivalence (parity) in the purchasing power of these two currencies. On the other hand, if from the original condition assumed, the American price level increased 100 per cent, British prices remaining the same, the rate would be £1 = $8.

The exchange rate tends to settle at purchasing power par for the following reason. Suppose the par is £1 = $4, but that the market price of sterling in New York is $5. In terms of what it will buy in England and in this country, sterling is overpriced in New York. Americans therefore would be loathe to part with $5 to obtain a pound sterling to purchase things in England when they could buy the same amount in the United States for $4. At the same time, Britons would be very willing to obtain $5 for a pound sterling and obtain 25 per cent more for their money in the United States than they could by spending at home. American imports from Britain would

fall off and exports to Britain would rise; therefore the demand for sterling would decline, the supply would increase, and the rate on sterling would begin its fall. When it had declined to $4, purchasing power parity would be re-established, American imports would increase, and American exports would decrease to a stable level. If the rate continued to fall to $3, the situation would be reversed. British goods would flow in large volume to the United States because of the low price of sterling. United States exports to Britain would decline because in London dollars would cost too much (only $3 for £1) in relation to their purchasing power.

Eventually, with relative prices in England and the United States assumed at the outset, the exchange rate would tend to settle at £1 = $4 as a result of these equilibrating influences. At this amount the rate of exchange would be "neutral," neither encouraging nor discouraging imports or exports in either country. Nor would the rate stimulate abnormal investments and movements of funds between the two countries or otherwise retard normal movements.

Purchasing-power parity is not to be regarded as wholly distinct from the gold standard; on the contrary, gold par represents purchasing-power parity, with the addition that *gold movements act to correct disparities in relative commodity prices, while keeping exchange rates within the gold points.* Similarly, governmental controls over exchange rates cannot ignore some rough parity of purchasing power, for otherwise the rates established would be difficult to maintain midst the inescapable forces of demand and supply. Any rate that expresses considerable disparity will, in the absence of governmental control, tend to change; if a disparate rate is established by control authorities it will result in financial difficulties and, eventually, further controls aimed to correct the disparity.

Stabilization Fund Control

Wide fluctuations in foreign exchange rates create uncertainty and risk in foreign trade. A considerable degree of stability, however, can be attained without gold movements by the actions of the so-called stabilization funds. In 1932 the government of Great Britain established its Exchange Equalization Account, and in 1934 the United States created its Stabilization Fund. The British fund was authorized to buy and sell foreign exchange, to buy and sell gold, and to hold balances or deposits in foreign banks. The United States fund had similar powers and was given the responsibility of maintaining the value of the dollar in the international market for exchange.

These large funds [10] were managed in co-operation with each other and with the financial measures of other foreign governments to prevent wide fluctuations in foreign exchange rates.

Suppose that the governments of the United States and Great Britain decided to keep the pound-dollar rates within a range of $4.47 to $4.53 per pound sterling. Their respective stabilization funds would then buy and sell exchange (and gold when necessary) to keep the rates within this range. Assume that there occurs a rise in the sterling rate in New York to the upper limit of $4.53, because the sterling balances of New York banks are low in London. The British fund may intervene by selling all the sterling exchange the New York banks want at $4.53. The rate in New York therefore would not rise above this figure.

The sale of sterling by the British fund is a purchase of dollars; therefore it acquires dollar deposits in the New York banks. If later the sterling rate in New York declines—or the dollar rate in London increases—to $4.47, the British fund may reverse its operations. If the decline occurs because American bank deposits in London are excessive because American banks have bought large quantities of sterling exchange from exporters, the fund will buy sterling bills in New York for no less than $4.47. It buys these with checks drawn on its New York bank deposits. The sale of sterling drafts by the New York banks to the British fund will mean a reduction in their deposits in London; the purchase of these drafts will in turn reduce the fund balances in the New York banks. If the British fund can buy enough sterling at the price of $4.47, the rate will not fall any lower. But suppose in buying sterling drafts that it has exhausted its dollar balances in New York? In this event, it could sell gold to the New York banks, replenish its dollar deposits in this way, and thereby continue to buy sterling at $4.47.

The American fund may operate in the same way. When in this instance sterling fell in New York (the dollar rose in London), it could buy sterling (sell dollars) at the $4.47 rate, and sell sterling (buy dollars) at the $4.53 rate. If it exhausted its sterling deposit in London, it could ship gold to London, build up its sterling balance, and thereby continue to purchase dollars (sell sterling). In this way, it would support the value of the dollar in the foreign market (or hold the sterling rate in New York down to $4.53). When both funds co-operate to hold the exchange rates within these limits, the prospects of success are increased.

The success of this method of exchange rate stabilization depends upon

[10] The United States Stabilization Fund had 2 billion dollars derived from the "profit" arising from devaluation of the dollar in 1934. The British fund obtained 575 million pounds (at that time roughly 2.5 billion dollars) from its government.

(1) agreement, by both stabilization funds, i.e., both governments, on the permissible range of rates, (2) the amount of resources available to the funds, and (3) the selection of the appropriate range of rates. The importance of agreement is clear enough. Also the need to have ample resources is evident, for these funds may be called upon to buy and sell enormous amounts of exchange. The particular range of rates affects the flow of payments between the countries. If the rates represent the approximate purchasing-power parity of the two moneys, the stabilization operations would be successful and the flow of payments would tend to balance without disturbing the monetary resources of either country. On the other hand, if the rates established represent the overpricing of sterling and therefore the underpricing of dollars, the balance of payments would be in favor of the United States and as a consequence the British fund would eventually become bankrupt, just as the overpricing of the dollar (underpricing of sterling) would eventually bankrupt the American fund.

This is a serious danger inherent in the stabilization-fund technique. It is not peculiar to this form of exchange stabilization, however. If gold pars diverge materially from purchasing-power parity, the country with overpriced currency would suffer a persistent loss of gold to the country with the underpriced currency. Similarly, if the rates established by the International Monetary Fund are out of line with purchasing-power parity, trouble will arise in the countries with overpriced currencies.

The British and American stabilization funds substantially lessened exchange fluctuations in the prewar years. Through agreements with France and with other countries the area of stabilization was gradually extended. Indeed, a few large funds of this type, working in harmony with each other and with the central banks of other countries, could actually establish a high degree of foreign exchange stabilization. This technique, however, was passed over in the postwar plans for exchange stabilization in favor of the far more comprehensive and complex International Monetary Fund.

The International Monetary Fund

The International Monetary Fund is a war-born international financial institution which was created by agreement of 44 countries in 1944, at the Bretton Woods Conference. It began operations early in 1947. Its broad purpose is to promote international financial stability and thereby aid establishment of an orderly flow of world trade. The creation of this fund was largely the work of the United States and Great Britain, as leaders in international finance.

In some respects the Monetary Fund is an outgrowth of the prewar stabilization funds, although vastly broadened in scope and operations. The main problems of the postwar world which fall within the orbit of the Fund are:

1. *Exchange-rate instability.* Prewar instability has been briefly described. The difficulties creating this instability were greatly aggravated by the disruptions and financial stresses of World War II. In the absence of some concerted effort to secure stability it was expected that postwar instability and exchange depreciation would be rampant throughout much of the world. The Monetary Fund is intended to prevent much financial disorder.
2. *Shortages and threatened shortages of foreign exchange.* The financial resources of many nations were depleted by the war, and the capacity of many nations to obtain foreign balances through exports was reduced or virtually destroyed by war damage and by internal economic and political disorganization. Many nations need foreign exchange balances to obtain imports vital to the revival of their industries. The Monetary Fund is a means of providing limited short-term loans to enable these countries to meet pressing needs for foreign exchange.

The International Monetary Fund has other corollary objectives: (1) providing short-term loans for countries experiencing balance of payments difficulties; (2) reducing restrictions on foreign exchange transactions; (3) recommending internal economic and financial policies to governments encountering difficulties; and (4) establishing a means of consultation and exchange of views and information on economic and financial problems. Some of these will be discussed briefly in the next chapter.

As an international organization the International Monetary Fund is independent of any national state, although the member nations have voting power and therefore joint control over its policies. This voting power is approximately proportional to the quota of money invested by each country. The management of the Fund is in the hands of officers and a directorate. The rights, privileges, and responsibilities of the Fund, and the member nations in it, are defined in the Articles of Agreement, the basic charter of the organization.

The financial resources of the International Monetary Fund are derived initially from the subscriptions of each member nation to its quota. These quotas were reached by common agreement of the members: the United States quota is 2.7 billion dollars, the United Kingdom's 1.3 billion dollars, and so on down to relatively small amounts established for the small

members.[11] The total quotas were expected to bring the resources of the Fund to more than 8 billion dollars, composed partly of gold with the remainder in the currencies of the member nations. Each country's quota represents its obligation to contribute to the Fund's resources; moreover, each country is entitled to borrow from the Fund an amount of foreign currency (foreign exchange) equal to its quota plus its gold subscription. The Fund is therefore an international lending agency, but with the stated provision that its loans are (1) for short periods of time, and (2) for emergency purposes only. It must be emphasized that the Fund is not intended to supplant the normal operations of the foreign-exchange market. As before, the banks of each country will hold balances in foreign countries, private traders will buy and sell exchange, gold movements will be made between countries, and credits or loans will be extended by banks in one country to the banks of another. The enormous resources of the International Monetary Fund will supplement these ordinary sources of foreign exchange. Some nations may never need to draw upon the Fund's resources; for example, it is unlikely that the United States will need to borrow foreign balances or foreign exchange from the Fund. The Fund merely provides each member with a "line of credit" which it can draw upon if its normal sources of foreign exchange have been temporarily exhausted.

The quota provisions are as follows: Each member must subscribe the smaller of the following amounts in gold: (1) 25 per cent of its quota, or (2) 10 per cent of its total gold reserves. If country A has a quota of 100 million dollars it would be required to subscribe 25 million dollars in gold; however, if its total gold reserve is only 125 million dollars it would be required to subscribe only 12.5 million dollars in gold. Its quota, plus the gold subscribed, measures the extent to which it may borrow from the Fund. The gold thus subscribed becomes the property of the Fund. In the course of time the Fund's gold reserves are expected to exceed 1.5 billion dollars. The gold subscription provision in the Agreement is a link between the Fund and the old gold standard. In one sense, the Fund has been established as a market for gold and thus gold maintains its pre-eminence as an international means of payment.

The balance of a country's quota is subscribed in its own currency or in its noninterest-bearing note. Country A therefore creates a deposit account in its central bank to the credit of the Fund, and the Fund draws upon this deposit as it needs country A funds. In the present illustration, the currency (deposit) subscription would amount to either 75 million dollars or to 87.5

[11] The Soviet Union had been assigned a quota of 1.2 billion dollars. However, as of the end of 1947, it had not become a member of the Fund.

million dollars (in country A's currency). If currency is not subscribed, country A gives the Fund a noninterest-bearing obligation which can be converted into a deposit account if needed. The nongold subscription to the Fund is thus a huge reservoir of balances of foreign money (foreign exchange) contributed by each member and available to be lent by the Fund.[12]

Suppose that country A needs foreign exchange to meet its current obligations, e.g., payment of imports and the service of its debt, but its normal sources of foreign exchange are exhausted. Unless it obtains additional foreign exchange, country A must curb imports and take steps to encourage exports, but the latter, especially, is apt to require some time for fruition. The Fund, however, stands ready to make temporary loans of foreign exchange to country A. The latter can borrow up to 25 per cent of its quota in any one year, but it cannot run its debt to the Fund beyond the amount of its quota plus its gold subscription. Interest is charged on these loans with the rate increasing as country A continues to borrow larger amounts. In this manner country A can obtain American dollars, South African pounds, Mexican pesos, and other exchange to pay its obligations. As it borrows foreign moneys, however, it must deposit an equal amount of its own money to the Fund's account in its national central bank. In effect, country A buys foreign money or exchange from the Fund on the security of its own money. The Fund reduces its holdings of dollars, pounds, and pesos and increases its holdings of country A's money.

Country A can reduce its loans from the Fund by (1) paying gold to the Fund, or (2) by building up its balances of dollars, pounds, and other money through increased exports and paying the Fund from these balances.

The Fund thus stands ready to make loans to a member experiencing an unfavorable balance of payments, or a "deficit" in its balance, when the member has no other resources to meet this problem. When its balance of payments becomes favorable the borrowing member is expected to repay its loan. The Fund as a short-term lending agency performs a function analogous to that of a Federal reserve bank which lends (makes advances) to a member bank temporarily in need of additional reserves. It must be emphasized that the Fund's function as a short-term lending agency is not intended to supplant the normal foreign-exchange operations, or the practice of short-term international lending by banks. It has been established to provide foreign exchange only in cases in which normal sources are inadequate and to prevent fluctuations in rates resulting from exchange shortages. Moreover, because it can lend only limited amounts, the Fund as constituted

[12] When Fund operations began in 1947 there were 39 member countries and therefore 39 different national currencies deposited to the credit of the Fund.

cannot play an important role in a major currency crisis, such as the "dollar shortage" problem of 1947. Had such a problem been foreseen the Fund might have been established with more flexible powers, permitting it to extend credit more freely than as originally provided.

The agreement establishing the Fund also provided for the stabilization of foreign-exchange rates. The currency of each member is defined in terms of gold and United States dollars. The gold content is the *par value* of each currency; the original par for each member is based on its exchange rate prevailing early in 1947. The gold content of the dollar, as established by Congress, is 0.888 grams of fine gold. The gold content of the pound sterling is established at 3.581+ grams of fine gold, because the dollar-pound rate early in 1947 was £1 = $4.03 (3.581 grams is 4.03 times the weight of the United States dollar). In like manner, the gold content of other member currencies has been established and, by the same token, their pars in relation to the dollar and to each other. Thus, in terms of the dollar, the Dutch guilder is 37.69 cents, the Honduran lempira is 50 cents, the Canadian dollar is 100 cents, etc. The actual rates of foreign exchange will vary above and below these pars by a small percentage depending upon the supply of and demand for exchange, but these variations presumably are to be kept within limits similar to those existing under the old gold standard.

It will be observed that the Fund Agreement follows closely the old gold standard in that all member currencies are (1) expressed in gold, and therefore (2) pars of exchange are expressed in gold. It departs from the gold standard, however, by allowing for modification in par values (gold contents of currency) in the following way. Any member may alter its par by as much as 10 per cent if it considers the original par to be disadvantageous. If the change proposed by a member is in excess of 10 per cent the Fund authorities may agree or object. Should a member alter its par value in spite of Fund objection, the member may be declared to be ineligible to use Fund resources and may be forced to withdraw from membership. On its own initiative the Fund may propose alteration of a member's par value. These provisions thus permit a measure of flexibility in foreign-exchange rates through the collaborative action of the Fund's membership and a departure from the rigidity of the old gold standard. It departs also from the unilateral decisions prevailing in a period (such as the 1930's) when each country altered its rates to its own advantage or in accordance with the exigencies which it had to face.

The International Monetary Fund therefore is (1) an international reservoir of financial resources for short-term loans to members in case of need, (2) a means of stabilizing foreign-exchange rates through a system of

gold pars, and (3) a means of collaborative and orderly modification of foreign-exchange rates in accordance with long-run economic conditions necessitating these changes. By the middle of 1947, the Fund had started its operations on a limited scale. As time goes on, it might become the decisive influence in establishing the structure of exchange rates for a large part of the trading world, just as the traditional gold standard was in earlier decades.

The foregoing discussion has stressed the determination of the exchange rates governing the great flows of money payments which comprise each nation's balance of payments. The rates determine the terms on which the currencies of the world are exchanged for each other and therefore the terms on which commodities and services are purchased and sold. They influence the short-term movements of credit (or bank deposits or balances) and gold. The flow of long-term investments or capital among countries is another important aspect of international finance which acts and interacts with these transactions.

LONG-TERM CAPITAL INVESTMENT

The purchases of foreign securities by United States nationals and the sale of securities to foreigners were shown in Table 75 and discussed as part of the international balance. In 1938 these purchases and sales were approximately equal; in 1946, however, the United States purchased *net* about 3 billion dollars of foreign securities. For many decades the purchase of foreign stocks and bonds by American investors has exceeded (with some exceptions during the 1930's) sales of securities to foreigners. During and after World War I, and again after World War II, United States government loans to European governments were substantial. This country has been, therefore, a capital-exporting nation for many years and bids fair to continue this role indefinitely into the future. Its capital exports or investments have made it a *creditor* nation; i.e., on balance its investments abroad are greater than the investments of foreign nationals in this country.

The principal industrial and commercial nations have long exported capital. During the nineteenth century British investments abroad were substantial, and Britain became the leading creditor nation of the world—a position which it finally relinquished because of the financial drain of World War II. Other western European countries as well as the United States vied with Great Britain in the international investment field, especially during the present century. The expansion of foreign investments by these countries lay behind the great growth in foreign trade in the past

century and the expansion of productive capacity throughout most of the world. Foreign investments in the past century were primarily made by private enterprisers in a country and were undertaken for profit. The private-enterprise system spread and grew throughout much of the world by means of foreign investments.

On the surface foreign investments are *financial* transactions. American investors purchase foreign stocks or bonds; the foreign company gets dollars on deposit in New York; in the course of time the American investor receives dividends or interest and eventually the bonds may be redeemed by the foreign company. These financial transactions are carried out in the capital market and foreign-exchange market. Underlying these foreign investments, however, are the physical movements of trade and services from one country to another. The United States has exported capital for many decades. How and why does it export capital? This country, as well as others, exports capital because its productive capacity enables it to sell more goods and services (and especially goods) abroad than it buys from abroad. It has been shown that this country had enjoyed a favorable balance in its merchandise trade for many decades.[13] The value of its exports has exceeded the value of goods imported each year by several hundred million and even by several billion dollars annually. This difference in the long run is made up by (1) gold imports and (2) loans to foreigners and other investments. In other words, foreigners sell gold and securities to the United States to make up the difference between this country's imports and exports of goods and services. Capital exports or foreign investments, therefore, rest essentially upon an excess of merchandise exports over merchandise imports.

The connection between foreign investment and merchandise exports may be brought out by the following illustration. A Brazilian railroad company decides to expand its operations. Suppose it is unable to obtain the type of rails and equipment it wants from Brazilian manufacturers. It therefore floats a bond issue in New York and temporarily deposits in a New York bank the funds obtained from American investors. Subsequently, the company orders the necessary rails and equipment from American producers, pays for them by check drawn on its New York account, and ships the merchandise to Brazil. The loan or investment in this case leads directly to an American export. Instead of buying American goods, however, the company may prefer to purchase British rails and equipment. In this case, it buys sterling drafts from a New York bank, transfers its funds to London, and buys the British goods. There is no American export to Brazil; however, the New York deposit in London had been created by

[13] See Table 70, p. 746.

American exports to Britain somewhat earlier. *The sale of sterling to the Brazilian company is possible because of the favorable balance of trade with Britain.* Suppose now that the Brazilian company did not want any imported rails and equipment; it wanted Brazilian money. With its bank balance it purchases cruzeiro exchange in New York and thereby transfers its funds to Rio; or it purchases sterling and with this buys cruzeiro exchange in London. In either case, this operation presupposes an antecedent favorable trade balance on the part of the United States either with Brazil or with Britain.

A capital-exporting country, therefore, is one that has a favorable balance of trade (or more broadly, a favorable balance of payments position). Indeed its export trade is partly financed by the loans it advances to foreign countries; without foreign investment its export trade would have to be reduced.

A capital-importing country, on the other hand, experiences an unfavorable balance of commodity trade or an excess of imports over exports. As already noted, it pays for this excess of imports either with gold or by borrowing the funds abroad. It does so because its domestic productive capacity is inadequate to produce enough (1) to satisfy domestic demands, and (2) to export as much as it imports. Given its productive capacity, it would not need to borrow from abroad if it were willing to reduce domestic consumption, or if it reduced its internal capital expansion, or both. In these ways it could limit its imports to the value of its exports and thus avoid going into debt to foreign investors. A poor country unable to obtain foreign capital, or unwilling to do so for political reasons, tends to remain poor. However, a country that cannot at the outset produce enough to allow internal capital expansion and an "adequate" level of consumption may attain both in the course of time if it attracts foreign capital. This occurred in the United States during the greater part of the nineteenth century. Foreign funds invested in this country together with domestic investments led to the rapid industrial expansion of that period; without investment from abroad this expansion would have been slower. Similarly, without foreign capital the economic reconstruction of wartime devastated countries must be deferred, until the productive resources of the regions gradually perform the task unaided.

Private investments abroad are made in expectation of interest and dividend payments; loans are advanced in expectation of eventual redemption. Capital-importing countries, therefore, assume obligations to pay their creditors. Eventually these income payments and the repayment of loans require large exports of goods by the *debtor nations*, for in this way they obtain the foreign balances or exchange with which to make payment.

This point is of great importance. Consider again the Brazilian railroad company as an illustration. Its bonds sold in New York require an annual interest payment; the company must therefore acquire the necessary amount of American money to pay its obligations. It does so by buying dollar drafts from its bank in Rio. However, the Rio bank can build up dollar balances in New York only by purchasing the dollar exchange of Brazilian exporters to the United States. Not only interest payments but also redemption of the bonds can be accomplished if the Rio bank has a sufficiently large balance in New York, i.e., if Brazilians export more to the United States over a period than they import. This principle applies generally to debtor countries: they must enjoy a favorable balance of commodity trade in order to build up foreign balances and meet their obligations.

Looking at the matter from the standpoint of the creditor nations, it is evident that they can receive payments in the course of time only by accepting an excess of imports over exports. By purchasing large imports they provide the debtor nations with money to pay their (the debtors') obligations. Great Britain, for example, has had an unfavorable balance of merchandise trade for many decades. The British have been able to purchase this excess of imports because of large interest and dividend payments received from foreign investments. At the same time, foreigners have been able to make these payments because the British bought their goods.

Financial Disruption

Foreign investment has been dealt hard blows during the last two decades. The depression of the 1930's made foreign lending difficult for the creditor nations; the subsequent decline in foreign trade, and especially in the exports of debtor nations, made the payment of interest and dividends difficult and often impossible. One after another, the nations of the world raised obstacles both to trade and to the flow of money. Each nation sought to restrict imports of goods, while many imposed curbs on payments to foreigners. The value of foreign investments dropped; some became worthless. In a few countries, such as Mexico, foreign capital was seized by governments and foreign owners were expropriated. Moreover, the problems of this period were magnified by World War II. The destruction of war created at once a great need for capital for reconstruction, but the capacity of the industrial nations of Europe was so reduced that it could not supply these needs. At present, not only Europe but also Latin America and Asia require foreign capital for expansion. Economic and political difficulties, however, stand in the way of private investment in many instances. Previous losses have

tended to deter additional private investment, especially if economic and political conditions are not currently propitious. Consequently, to provide a source of long-term capital that might otherwise not be available, the International Bank for Reconstruction and Development was established by the Bretton Woods Agreement.

The International Bank for Reconstruction and Development

Like the International Monetary Fund, the International Bank is a co-operative financial institution with capital subscribed by the member nations. Each member is assigned a quota or capital subscription; total subscriptions are expected to approximate 9 billion dollars, of which one third is the United States quota. Initially, only 10 per cent of each member nation's quota must be paid into the Bank.

In its lending activities the Bank is empowered to do the following:

1. It may lend directly to governments or enterprises from its capital funds. Loans of this type outstanding cannot exceed 20 per cent of its capital.
2. It may borrow funds by selling its own securities in the capital markets, and may lend these funds directly to governments and enterprises.
3. It may guarantee loans made by foreign borrowers in the private capital market.

The plans which created the Bank stressed the third type of financing over the direct loan. Under the guarantee provision, a borrower in a foreign country sells bonds to private investors in New York, and the Bank guarantees that both interest and principal shall be paid in accordance with the loan contract. The advantage of this guarantee to the investor is obvious. The borrower derives an advantage because, without the Bank guarantee of his loan, he might be unable to sell his bonds, or might be able to do so only under onerous conditions. In return for this guarantee, borrowers are obliged to pay an "insurance premium" or commission to the Bank. These funds go into a pool administered by the Bank; if necessary they may be used to protect investors against loss due to default of the borrower. In a sense, the Bank is a mutual loan insurance system; it guarantees loans in a way similar to the guarantee of home mortgages by the Federal Housing Administration.

Loans made or guaranteed by the International Bank are backed up (1) by the assets of the borrower, (2) by his government, and (3) by the resources of the Bank itself. Thus investment risk is minimized so far as the

lender is concerned. Because of the Bank's responsibility as guarantor, the Bank is required to investigate loan applications carefully and to approve only those that contribute to the productive activities of the borrowing country.

The Bank is not intended to supplant private lending along traditional lines. A country can avail itself of the Bank's facilities and resources only if other sources of foreign capital are not available, or if the terms of private loans are burdensome. It is possible, however, that the Bank will tend to influence the rates of interest and the terms of foreign lending in its capacity as an actual or potential competitor in the field of international capital movements.

It is too early to assess the influences of the International Bank and the International Monetary Fund on international finance. Their combined functions embrace exchange-rate stabilization, the promotion of free multilateral trade, the provision of short-term capital, and the creation of a mechanism to provide long-term capital to nations in need of reconstruction and expansion. Their purpose generally is to establish sound international financial conditions for the encouragement of world trade and the growth of national income and output in the countries of the world.

CHAPTER 36

International Economic Policy

Governmental policy plays an important part in the flow of goods and funds among the nations of the world. Since each country has a stake in foreign commerce, each government seeks to establish salutary conditions for the conduct of its trade and financial transactions. Mainly because of different political and social precepts but also because of varying amounts of economic resources, governmental policies may encompass a wide range: from free trade on one extreme, to complete governmental control of all aspects of trade and finance on the other. In decades past Great Britain was the classic example of free trade, while at present the Soviet Union

with its governmental monopoly of foreign trade represents the fullest measure of control by government.

In the modern period most national governments have increasingly intervened in the regulation of the volume, types, and direction of foreign trade. This interventionism has paralleled the rise of control of domestic economic activity by governments, i.e., in those countries—mainly Western—which previously followed a *laissez-faire* domestic policy. Today, the financing of trade, i.e., the acquisition and payment of foreign exchange, and the volume and types of imports and exports are more or less rigidly controlled in all but a few countries of the world. Since in general governmental policies are restrictive, international trade has become burdened with regulations and prohibitions. The United States now imposes fewer controls than almost any other country; yet the policy of this country is a far step from freedom of trade. Normally few burdens are placed on exports by this country (the Constitution forbids export taxes), but tariff restrictions are imposed on imports, which of course reduce other countries' exports. Some countries impose controls over other aspects of foreign exchange as well as domestic production, so that not only the amount and direction but also the composition of trade are affected by government policies.

It is readily evident that a restriction of trade by one country entails a forced restriction on another, since international trade is a process of exchange: no country can limit its imports or exports directly or indirectly without diminishing the total amount and character of international trade. But this judgment presupposes a condition of unfettered or free trade. It is difficult not to regard restriction of trade as a departure from a "natural" state of free trade, even though free trade is not obviously or undisputably the ideal requisite for commerce between nations. Nevertheless, the fact that before 1914 trade was conducted with only minor restrictions, and the possibility that present-day barriers may be removed, raise many important questions. What is the effect of restrictive policies on world production and world trade? What effect would free or at least freer trade have on the use of resources and levels of consumption in individual countries?

Free-Trade Policy

Free trade is a policy of nonintervention by government in the foreign trade activities of its nationals.[1] Goods may be imported and exported with-

[1] A free-trade policy does not mean literally the absence of governmental obstacles. The imposition of standards on the import of goods for the protection of the health of the population is consistent with this policy. Similarly a tax (tariff) on imported goods designed only to raise tax receipts is not an abandonment of a free-trade policy.

out restrictions, and private persons and business firms may freely buy and sell their own and foreign currencies in the foreign exchange market. From the mid-nineteenth century until World War I, Great Britain followed a free-trade policy, and other nations of Europe, especially France, approached this policy during the third quarter of the nineteenth century. Moreover, the predominant British influence in international transactions established relatively free trade in many other parts of the world during this period. Since World War I, Great Britain has followed the interventionist path, at first cautiously, but as the emergencies of depression and a second world war arose, it became strongly restrictive. As a policy, free trade stands small chance of being revived in the calculable future either in Britain or elsewhere, but the influence of the free-trade argument and certain advantages of free-trade policy are nevertheless evident in the postwar proposals for international economic collaboration among nations.

The basic argument for free trade rests upon the economic advantage of maximizing the division of labor. If all countries allowed unrestricted imports and exports, productive effort would tend to be specialized throughout the world in accordance with the principle of comparative advantage. Each nation would then devote its resources to the most advantageous uses and would sell (export) the surpluses thus produced for goods produced most advantageously by other countries. It would therefore obtain foreign goods at the lowest possible price, or at the least expenditure of labor, capital, and natural resources. Advocates of free trade often recount the advantages of internal free trade in the United States. Automobiles produced in Michigan, cigarettes in North Carolina, women's dresses in New York, and movies in California may be shipped anywhere in the country without payment of tariff duties at state lines; and no state imposes a quota limit on the amounts received from other states.[2] Because the nation is in fact a single large market, production of commodities and services is highly specialized. The advantages of specialization are seen in low unit costs or high productivity. If automobiles, cigarettes, women's dresses, and movies were produced in all or most of the states, because of tariff or quota restrictions designed to encourage local production, specialization would be limited, resources would not be concentrated in their most efficient uses, the advantages of large-scale, low-cost production would be lost, and prices would be relatively high.

The economic argument for free trade proposes the application of full

[2] During the 1930's, however, the states did impose barriers to imports of certain kinds from other states. Such barriers, however, affect only a few commodities, and usually exist under the guise of regulation for some other purpose, such as sanitation.

specialization and thereby the most efficient allocation and use of economic resources between nations. Free trade would extend to a world context the advantages of an unfettered market, multilateral trade, and the operation of price-cost relationships which are realized within each country. The clarity of the economic argument, however, has availed little in present-day international affairs. In a world given over to the furtherance of a nationalism more intense and organized than that of the mercantilist era, free trade is a lost cause. The quest for national power, for protection of special interests, for insulation against foreign economic disturbances, and for collectivist economic planning—all nationalist aims—spell defeat for the free-trade goal.

GOVERNMENT CONTROLS OF FOREIGN TRADE

The Protective Tariff

A protective tariff is a tax or duty levied on designated types of foreign goods entering a country. The purpose of the tax is to raise the price of foreign goods sold in the domestic market, and thus to enable domestic producers of the same or similar goods to compete on advantageous terms with the foreign producers. The increased price—from the addition of duties—of the foreign goods usually lessens their sale; the volume of imports is reduced. Thus a tariff on sugar raises the price of foreign sugar; the higher price enables domestic producers to expand production or to maintain their volume of production at higher profits. The tariff therefore protects a domestic market and a group of domestic producers against lower-priced foreign goods; but for this protection some domestic industries would not exist at all, or at least would operate on a smaller scale.

Tariffs or import duties fall into three categories. The *ad valorem duty* is expressed as a percentage of the price of the imported commodity; e.g., the duty on cotton piece goods may be 20 per cent added to the price of the products upon entering the country. A *specific duty* is a levy of so many cents or dollars per physical unit of product; e.g., wool imported into the United States has been subject to a 34-cent per pound duty. A foreign product may be subject to both specific and ad valorem duties. This *combined duty* may be so many cents per pound or a certain percentage of the imported price, whichever yields the lower tax.

Foreign goods subject to import or customs duties are termed *dutiable imports;* those not taxed are termed *free imports.* Many goods enter this country as free imports, being in general of types not produced in this

country. On the other hand, dutiable goods are usually those which are the same as, or similar to, goods produced in this country.

For many decades the United States has followed a highly protective tariff policy. Early in the nineteenth century the small manufacturing industries of this country, many of which were the first producers in new fields, sought and obtained a measure of tariff protection from Congress. During and after the War between the States, tariff barriers were raised to high levels and were extended to an ever-increasing number of products. The important shift to high protective duties during this period was directly related to the change in political power resulting from the war. Before the war the South as a political bloc resisted protective tariffs in most instances. Its cotton and tobacco needed no protection from foreign competitors, for the South excelled the world in the production of these goods. Moreover, it resisted protection for Northern manufacturers because import duties would have (1) increased the prices of both Northern and imported manufactured goods sold in the South, and (2) curtailed imports from abroad and thus limited foreign capacity to purchase Southern exports. The absence of Southern delegations from Congress during the war and early reconstruction periods removed their restraining influence. This, together with the increased power of manufacturing interests in politics, led Congress to establish protectionism as a national policy. The return of Southern political influence in the 1870's was powerless to reverse the movement, even though it continued to oppose tariffs and was later joined with other sections of the country in a liberal program. The Republican party became increasingly associated with industrial interests fostering protectionism, and thus was (and is today) known as the "high-tariff" party. Since the Democratic party sponsored low tariffs (but not wholehearted free trade), tariff policy became a political issue between the two parties. Although the issue remains, its political effect and the party distinctions are no longer so important.

The case for protectionism is made up of many arguments, of varying validity.

1. *A small industry is at a disadvantage in competing with a large, well-established foreign industry because the latter enjoys economies of large-scale operations which the former does not.* Without tariff protection the small industry, with its relatively high unit costs of production, could not survive the flood of imports coming from the low-cost foreign country. Tariff protection increases the price of foreign goods in the domestic market, thus enabling the small domestic industry to survive and grow. In the course of time the small industry becomes

large and efficient and is fully able to compete with foreign producers. At this point protection can be withdrawn, i.e., the commodity or commodities produced can be removed from the dutiable list. This "infant industry" argument was widely used by manufacturers in the early stages of American industrial life, as it is today by some Latin-American countries that are attempting to industrialize in the face of competition from foreign industries.

The argument is valid when it is used to support a tariff for a small industry which clearly has good prospects of reducing unit costs by large-scale output; but in practice, too often it is used to justify tariff protection for industries whose prospects of success are dubious. Moreover, experience in this country shows that protection once granted is not easily withdrawn; the erstwhile "infant" clamors for protection perhaps for other reasons and may be successful in preventing a real test of its ability to compete.

2. *Tariff protection enables the protected industries to grow by taking over all or part of the domestic market previously supplied by foreign producers.* In this way national output as a whole increases. At best, this argument has limited validity when the *short-term* effects of tariff protection alone are considered. It is possible that an increase in tariff duties would encourage industrial expansion in the protected industries; however, the ensuing reduction of imports would before long cause a reduction in exports as well, with depressing effects on export industries. Thus the expansion of one group of industries might be offset somewhat later by a decline in other industries; tariff protection directly encourages a less efficient industry and indirectly discourages a more efficient industry. Resource allocation, therefore, is changed at the expense of productivity. It is conceivable that total employment may be increased in this way, but it is also true that total employment may be increased by reverting to the pick and shovel in the construction of highways.
3. *The purchase of foreign goods takes money out of the country; on the other hand, the purchase of domestic goods keeps money at home.* This "buy-at-home" argument is hoary with age, being often used locally to discourage consumers in one town from buying in another town. Literally the logic leads to a negation of trade of any sort. Why should not a family make its own clothes, do its own baking and laundry, build its own home? The speciousness of this argument is evidenced by the fact that a nation's imports provide the foreigner with money to buy that nation's goods. The dollars used to buy foreign

goods are spent by the foreign country directly or indirectly to buy American exports. In support of tariff or other restrictions on imports the keep-money-at-home argument can be rejected as unsound and even naïve. Under some circumstances, however, a similar argument has short-run validity. If a country imports heavily and suffers an adverse balance of payments its foreign exchange balances will be depleted. A reduction of imports might be necessary in order to conserve these balances or to prevent a loss of gold. But such a course is not actually based on the buy-at-home position.

4. *Tariffs protect the high wage rates and the high standard of living of an advanced country against the "pauper" labor of poorer foreign regions.* Proponents of protective tariffs have exhibited considerable solicitude for the American wage earner in advancing this idea. As a general argument it is specious. Hourly wage rates in American industry in general are far above similar rates elsewhere, but high wages are not to be confused with labor costs. The typical American wage earner in manufacturing industries works with more efficient capital facilities and therefore produces more goods per hour or day than do wage earners abroad. As a consequence, *labor costs per unit of goods* may be less in American industry even though hourly wage rates are considerably higher. Indeed, this wage argument has been made to work both ways; Americans have argued for tariffs to protect the American workman from the low-paid labor abroad, and foreigners have argued for tariffs to protect their labor from the more efficient, high-paid labor of the United States.

   Under some circumstances, however, high wages may mean costs in excess of those abroad. If an industry abroad has similar capital facilities and enjoys economies in production comparable to those in the same industry in the United States, then its lower wage would enable it to undersell its American competitor. In the absence of tariff protection the American industry would find itself in difficulties; its unit labor costs would exceed foreign labor costs. This, however, is not typical of the major productive activities involved in the international market. In short, the tariff has little or nothing to do with determining or maintaining wage rates; wage rates are governed largely by productivity.

5. *Tariff policy can be designed to equalize production costs between domestic and foreign industries.* The so-called "scientific" tariff was intended to achieve this equalization. The United States Tariff Commission many years ago was charged with the responsibility

of measuring domestic and foreign costs and recommending tariff rates based on cost studies for Congressional consideration. The measurement of costs, however, is a formidable task. Which costs are to be taken? Those of the most efficient, average, or least efficient firm in an industry? Which accounting system should be used? What can be done about an industry whose firms keep only the most rudimentary financial records, or none at all? Even if a method for measuring costs of production for American firms was agreed upon, how would costs in many foreign countries be determined? The difficulties suggested by these questions have cast doubt on the cost equalization basis for tariff rates. Not only is the method probably impracticable, but the goal—if undertaken to be realized completely—is contrary to the very existence of trade, for if costs are actually equalized, trade would cease.

6. *Tariffs are protection against dumping by foreign producers and governments.* Dumping is the sale of foreign goods in a country at prices ruinously low for the domestic producers. It is a practice followed occasionally to get rid of large surpluses of goods. Sometimes it is predatory: producers in one country dump products in another for the purpose of ruining producers in the importing country. In the early 1930's the Soviet government sold substantial amounts of timber in European countries at prices far below those of lumber producers in those countries. This sale might have been intended to cause economic difficulties in the importing countries, or, more probably, the Soviet government needed foreign exchange quickly and used dumping tactics to obtain it. The tariff is a means of protection against dumping which many countries have used to meet this problem.
7. *Tariffs give to a nation the advantages of economic diversification.* This argument, unlike most of those described herein, is usually advanced in support of broad national purposes rather than for special interests. Briefly the position is as follows: a nation should guide its productive activities so as to attain a balanced production and a high degree of self-sufficiency. Protective tariffs are means of encouraging industries that otherwise would not exist or that would not be able to attain the desired size because of foreign competition. The advantages of diversification are several from this point of view.
    a. *Military security.* Many governments have protected industries essential for war purposes. European governments for decades have protected agriculture by tariffs to attain self-sufficiency in food production. Chemical, steel, and other strategic industries

have been similarly protected. Britain's abandonment of free trade following World War I took the form of protecting "key" industries, such as chemicals. Adam Smith, the great spokesman for free trade, admitted military security as the important qualification to his plea for economic liberalism.

b. *Economic stability.* A wider and more diversified economy is believed to be less likely to experience a great degree of economic instability. Countries dependent upon a few main export items, such as raw materials, are susceptible to great instability; when foreign demand for their products falls off, these countries are depressed. Therefore, many raw-material-producing countries have encouraged the expansion of manufacturing by tariffs, subsidies, and other means, and industrial countries have favored certain areas or types of agricultural production.

c. *Economic nationalism.* Closely allied to both these advantages is the desire of nations hitherto "undeveloped" to become economically important. In part, this is a reaction against dependence upon the industrial nations for manufactured goods, i.e., against the "colonial" status this dependence engenders. In recent years economic nationalism has become a crucial issue in some nations and has led to intensified "five-year plans" designed to place the maximum of national resources into building up industrial capacity. Political leaders in many of these countries enshrine industrialism itself as a national goal and measure national greatness in terms of pig-iron capacity—a criterion imported in part from the United States.

Some General Considerations

The case for protective tariffs could be filled out in more detail but the arguments given describe it sufficiently. The United States has been and is a stronghold of protectionism, chiefly because producer interests have been influential and sometimes dominant in national policy. American tariff policy has been based on special pleading and implemented by political pressure tactics for the most part; business generally has persuaded itself and the public at large that this form of governmental intervention is in the national interest. The Hawley-Smoot tariff of 1930 represented the high-water mark of protectionism in this country; since then some reductions have occurred under the Reciprocal Trade Agreements Act. Despite these, however, the American tariff remains high.

The protective tariff is clearly an obstacle to foreign trade. Imports into this country undoubtedly would have been larger had this country followed a more moderate tariff policy. American tariffs have, moreover, led foreign governments in turn to raise tariff barriers against American goods as a retaliatory measure, or as a means of conserving their foreign exchange balances. "Tariff wars" or tariff retaliation excessively burden foreign trade and cause a relatively inefficient use of economic resources in the countries protected by tariffs.

Some detrimental consequences have resulted from the high-tariff policy of the United States, both for itself and for other nations of the world. America's position as a creditor nation was previously discussed. A creditor nation expects payment from abroad in the form of interest, dividends, and redemption of loans; however, debtor nations must, to make these payments, export commodities and services. Therefore, the creditor position of the United States should have involved the importation of large amounts of foreign goods, but its high-tariff policy has made this impossible. In the course of time foreign debtors encountered increasing difficulties in obtaining dollar exchange to pay American investors and to buy American goods. The wholesale repudiation of debts owed the United States during the 1930's may be traced in good part to the high-tariff policy of this country. This country was (and is) a creditor nation but has followed the foreign trade policy of a debtor nation.

During the 1930's some steps were taken to modify the high-tariff policy. The Reciprocal Trade Agreements Act of 1934 authorized the President to negotiate reciprocal trade agreements with other countries in order to foster foreign commerce. In return for tariff reductions and other concessions from foreign governments, the President was empowered to reduce import duties (on the products involved in negotiation) by as much as 50 per cent. Trade agreements were concluded with a number of countries and tariff duties were reduced here and abroad. World War II disrupted trade so drastically that the effectiveness of these modest tariff reductions can hardly be measured at this time. Nevertheless, it is indisputable that the act was a departure from the exaggerated protectionism of the 1920's. With the close of World War II the urgent need for a revival of world trade spurred efforts to reduce tariffs and other obstacles. The Bretton Woods Agreement and the international financial institutions provided under its terms have this as one important purpose. The Geneva Conference on Trade and Employment of 1947 was called for the purpose of bringing about a world-wide reduction of tariffs and other trade obstacles in a program to increase trade, production, and employment throughout the world.

The case for tariff reduction is persuasive. But a movement toward freer trade (not free trade) will encounter many difficulties. In each country are to be found protected industries which will be damaged by tariff reduction, and these interests will oppose tariff reform. More importantly, perhaps, most of the countries of the world fear the loss of gold or foreign exchange balances that might result from a substantial relaxation of import restrictions. Others have internal economic plans which require foreign trade controls, and some conduct foreign trade as a virtual government monopoly. The economic advantages of freer trade—i.e., a more efficient use of economic resources and consequent increased productivity—may conflict with internal political aims, conceptions of economic stability, and the entrenched interests who profit through protection.

Quota Restrictions

Many countries employ import quotas as a means of restriction. In general the quota involves a *quantitative* limitation on the amounts of goods that may be imported. In this respect it is a more effective control than the tariff, for the tariff merely imposes a tax (or increased price) on imports. Quota limitations were adopted by some nations during the 1930's to protect domestic producers and to conserve foreign exchange balances. There are several types of import quotas:

1. *The Tariff Quota.* This type combines a tariff and a quota. A specified amount of an imported commodity is permitted to come into a country at a certain import duty, but any additional amounts are subject to a higher duty. For example, the United States permits the importation of 22 million pounds of Cuban cigar-filler tobacco at an import duty of 14 cents per pound. Amounts imported in excess of this quota are taxed at 28 cents per pound. It will be noted that the tariff quota is not strictly a quantitative control, although the increase in the import duty beyond the allowable quota works in this direction.
2. *Fixed Quotas.* This type rigidly limits the amounts of a commodity (by value, weight, or number of units) that may be imported. During the 1930's France instituted fixed quotas on a large number of commodities and many other countries followed suit. This type of control has not been a part of normal American trade policy although during World War II import quotas (or allocations) were imposed on many imported commodities.
3. *Global Quota.* Quota controls in the past have varied considerably.

In some cases a country, by law or decree, imposed a limit on the importation of a commodity without reference to the amounts that came from individual supplying countries. This has been termed the global quota. Its defect lies in the possibility that one of the several supplying countries might ship all that is allowable under the quota, leaving the others out of the market entirely.

4. *Allocated Quota.* To avoid possible discrimination under the global quota, allocated quotas often are established so as to apportion the imported commodity among the several supplying countries.

The determination of quotas by the government of the importing country is a complex task. If a restricted commodity is essential (e.g., steel) to the importing country's productive process, care must be taken to ensure that at least an adequate amount is received. Moreover, it is necessary to assure that the limited amount imported is used most advantageously, and this requires rationing the sale of the imported commodity. Thus import control leads to a series of internal controls. On the other hand, if quotas are imposed only on "nonessential" imports the problem of scarcity may be handled adequately by market price.

In order to assure the limitation of imports within the quota, and to divide up the import business among the various trading firms, governments usually resort to *import licenses.* Each importer dealing in the particular trade under control is granted a license by his government which permits him to import a stated share of the total quota. Needless to say, this procedure is awkward and burdensome to the importer but effective in limiting imports.

In some instances fixed quotas are not imposed but all imports are subject to import license control. Importers must obtain a license from a governmental agency; licenses are issued in accordance with a planned import policy or in accordance with the size of the foreign exchange balance. This import license procedure is more flexible than one limited by a quota system, but it may also be more restrictive.

Foreign Exchange Control

Of all methods of regulating and restricting private foreign trade none is more effective than exchange control by governmental authorities. This control involves regulation of the purchase (acquisition) and sale (disposal) of foreign currencies. The most common type of exchange control requires (1) exporters to turn over to a governmental agency the foreign exchange

accruing to them from their sale of goods abroad, and (2) the agency to sell this exchange to importers and others in accordance with governmentally established criteria. The exporter thereby receives domestic money for his foreign exchange; in effect he is required to sell his foreign money to the government at a rate determined by his government. Importers must buy exchange from government as the only source, since all foreign balances become property of the government and can be used only for purposes designated by it. In this way government establishes complete control over the amounts and types of imports, foreign investments, and other uses of the country's foreign balances.

In practice, exchange control has exhibited many variations from the standard type described here. The rates at which a government will buy exchange from its exporters may vary; higher rates are paid for exchange on exports to certain countries in order to encourage trade with these countries. On the other hand, the rates at which a government sells exchange to importers are differentiated in an effort to encourage imports from certain sources and discourage them from other sources. Or the rate may be relatively lower for the purchase of some commodities in order to encourage their importation. Varying rates of exchange are possible under a system of exchange controls because of the monopolistic position of the governmental agency in the supply of foreign balances.

The multiplicity of exchange control practices may be included in several general categories, three of which require intergovernmental agreements and the fourth of which is unilateral: (1) *compensation arrangements,* (2) *clearing agreements,* (3) *payments agreements,* and (4) *unilateral control.*

A compensation agreement is essentially a barter deal. Country A agrees with country B to permit their respective firms to exchange certain amounts of commodities of equivalent value, the equivalence being determined at rates of exchange agreed upon by the two governments. Thus a firm in country A ships a quantity of chemicals to country B, and a firm in country B ships raw materials to A, both amounts of equal value as determined by the agreed-upon rates of exchange. Arrangements of this type were made by Germany and some other countries during the early 1930's because exchange shortages and depression difficulties made normal trade virtually impossible.

Clearing agreements were widespread during the prewar decade. For example, Germany and Switzerland agreed to pay their own exporters for the goods exported to the other country. A German exporter of coal to Switzerland was paid by the Reichsbank, while a Swiss exporter of watches to Germany was paid by the Swiss National Bank. Under the agreement

both banks handled the transactions at an agreed-upon rate of exchange; at the end of the clearing period one paid the other whatever difference occurred in the value of goods exchanged. Or, as frequently happened, Germany would build up a debt to its partner in the clearing agreement.

Payments agreements are intergovernmental arrangements under which two countries, one of them a debtor to the other, allocate foreign exchange. The Anglo-Argentine agreement of 1933 is an example. Argentine balances accruing in Britain were made available for payments on Argentine debts to Britain, for certain classes of British exports to Argentina, and for other agreed-upon uses. Although they resemble clearing agreements, payments agreements differ in that actual foreign exchange transactions continue, but within the framework of the agreement.

Unilateral exchange control involves no agreements between nations. A country exercising control merely limits the purchase of foreign exchange by its importers and other nationals having payments to make abroad. There need be no discrimination on the part of the government, favoring one country against another. Chile employed this technique during the prewar decade. However, to encourage the importation of certain goods at the expense of others, Chilean authorities established several exchange rates that favored the purchase of goods most desired for its internal economy.

Why were these various methods of exchange control adopted? Foremost among the reasons was the actual or threatened loss of foreign exchange balances during the international crisis of the early 1930's. As soon as financial difficulties arose, many nationals in the threatened countries sought to buy dollar or sterling exchange; i.e., they attempted to transfer their money to New York or London banks. Faced with this flight of funds, some countries imposed restrictions on foreign exchange transactions. Others waited too long and lost most of their foreign balances. Controls to prevent the flight of capital almost necessarily led to general exchange control, for it is difficult to distinguish between a purchase of foreign exchange to import goods and a purchase for safe-keeping abroad. In some countries exchange control came to be regarded as a means of collecting revenue. In Chile the government's buying and selling rate margins yielded a profit in excess of 25 per cent. Exchange control, however, was employed most widely as a means of effecting bilateral trade arrangements between countries. Germany employed exchange control as a means of channelizing trade, and in this way exercised considerable control over countries (notably the Balkan countries) dependent upon Germany as a market and as a source of goods. Others sought to channelize trade by restricting their im-

ports largely to their best customers for exports, or they sought to curtail imports from countries with which they had an unfavorable balance of payments. Controls over exchange transactions in most countries are added to the barriers imposed by tariffs and sometimes both of these have been combined with import quotas.

Exchange controls born of prewar financial exigencies and power politics persist in an aggravated and more comprehensive form in the postwar period. Early in 1947 all but eight nations had some form of exchange control. The United States is the only major country not using this device either to protect its foreign balances or to discriminate in trade among nations, and it may be said that this liberalism is made possible by its position as the leading producer and largest holder of gold. Because of the status of the United States in world trade, postwar use of exchange control is primarily designed to conserve dollar balances held by foreign nations. For it is hardly an exaggeration to say that, if the purchase of dollar exchange were freely permitted in a great many of the countries abroad, their nationals would transfer their wealth into American deposits and quickly strip the countries of the dollar exchange available for imports of American goods. Other reasons include the protection of domestic industries against foreign imports, the bilateral balancing of trade, and the control over foreign trade to assure the success of domestically planned economic systems.

International Commodity Agreements

Governmental intervention in foreign trade takes still another form in international commodity agreements. These are agreements among the governments of important producers and consumers of certain basic products. The International Wheat Agreement of 1942 was a contract between the principal wheat exporting nations—Canada, Argentina, Australia, and the United States—and Great Britain, a leading importer of wheat. The agreement was not put into effect because of war conditions; however, its provisions indicate the scope of agreements of this type:

1. *Export control.* Of the total amount of wheat and flour exported by the four countries the percentage share was to be as follows: Canada 40, Argentina 25, Australia 19, the United States 16.
2. *Production control.* The four countries pledged to control their domestic production (by whatever means they each deemed feasible)

in order to limit output to (a) domestic requirements, (b) their basic export quotas, and (c) additions to reserve stocks.

3. *Reserve stocks.* Each of the producing countries agreed to maintain reserve stocks of wheat on hand at the end of each crop year. The size of the reserve was established for each country. Should stocks on hand exceed the agreed-upon reserve in any or all countries, the excess would be disposed of domestically as livestock feed or other lower order uses.
4. *Prices.* Each year the countries to the agreement were to establish the world export price of wheat. These annual prices were to be "reasonably remunerative" to producers, "fair" to consumers in the importing countries, and in "reasonable relationship" to other prices. (These vague criteria relating to the determination of price leave this problem open to constant negotiation and friction.)

The point of importance for present purposes is the extent of trade control by governments to such an agreement besides the internal controls over production necessary to make agreements effective. Agreements by other groups of countries exist, or have existed, or are in process of negotiation, covering coffee, sugar, cotton, beef, tea, and rice. These products embrace a substantial part of total world trade. The departure from free, or relatively free, trade in the world market is dramatically revealed in this type of agreement, for the main elements in the trading process—production, exports, and prices—are governmentally determined.

International Cartels

A cartel is an association of, or merely an agreement among, producers in an industry for the purpose of regulating production, establishing prices, dividing up the market, and other monopolistic devices. Domestic cartels are common in Europe. The British steel and chemical industries and a major part of German industry before the war had been cartelized. Domestic cartel arrangements are contrary to the Sherman Antitrust Act in this country but somewhat similar results are achieved by the formation of huge corporate structures. Cartels extend into the international field when producers of the same products in two or more countries reach agreements on the division of the international market, export prices, the exchange of patents, and other aspects of international trade. As a rule, international cartels are sponsored and regulated by governments abroad. Thus the British government exercises a measure of control over such cartels as

the Imperial Chemical Industries, and the German government before the war directed many of the activities of I. G. Farbenindustrie. These huge private-industry groups thus become instruments of national policy and the international agreements made by these groups often constitute a kind of intergovernmental collaboration.

Several types of cartel organization may be distinguished. The *association* is a form similar to the trade association found in domestic industries. The international cartel association is composed of producers of a commodity or line of commodities from several different countries. Formal agreements may be made by the association to fix prices, control production, regulate exports, allocate markets, redistribute profits, and effect an exchange of patents among the members, and to establish a combined sales organization to handle all export trade. American firms legally can participate indirectly in these associations through export associations provided for by the Webb-Pomerene Act. Secret participation on a doubtful legal basis is, however, practiced by some American firms.

The *patent-licensing* agreement is another form of international cartel; it is most common in the petroleum, chemical, and electrical industries. Firms in possession of patents license the use of these patents in other countries; the licensee may be required to limit his output, exports, fix prices, and follow other policies established by the cartel.

A third form of international cartel is the *combine*. For example, the du Pont company in this country and Imperial Chemical Industries in Britain formed the Duperial companies in Brazil and Argentina. The parent companies conduct business in these countries only through the Duperial subsidiaries; their subsidiaries manufacture goods, import du Pont and Imperial Chemical products, and manage the sale of these products in their areas. The effect is that competition between du Pont and Imperial Chemical is avoided by a division of the market. The holding-company structure is also used to effect cartelization. The International Telephone and Telegraph Company owns or controls cable, radio, telephone, and equipment companies in many different countries. Under contractual arrangements with the Dutch holding company, N. V. Philips, producer of electrical equipment in many countries, these two parent organizations can achieve a high degree of concerted business policy throughout a large part of the world.

Cartel arrangements reduce competition in world markets in many ways. The United States Department of Justice listed 179 cartel agreements in effect in 1939, with American participation in more than 100. By these arrangements the volume, direction, and types of world trade are subject

to control, as are the prices of the controlled commodities. It should not be inferred that cartelization necessarily reduces production and exports; like some domestic monopolies these international agreements in some instances may encourage output. On the whole, however, the evidence points to restriction and to monopoly pricing.

Since many international cartels are governmentally sponsored and regulated, their policies in effect are quasi-governmental. Cartel arrangements are in some instances akin to international commodity agreements made by governments in that both allocate export markets, control prices, and regulate production. Cartel agreements are "private" while governmental commodity agreements are "public," but the frequent governmental participation in cartels as well as the personal interests fostered by international commodity agreements serve to weaken this distinction.

Governmental Trading Monopolies

Governmental control of foreign trade reaches its apogee in the collectivist regime of the Soviet Union where importing and exporting are governmental monopolies. The socializing of industry in other countries, especially in Europe, will make "state trading" a growing factor in future international economic policies. Under the totalitarian economic system of Russia, the state determines the amounts and types of imports needed, as well as the exports required to secure the necessary foreign exchange. Tariff protection, quotas, and exchange controls are wholly unnecessary in the conduct of its international economic policies. Foreign goods cannot compete with and threaten Soviet production, for the government does all the buying from abroad: the state monopoly achieves directly and completely the power which is in part sought by other countries using restrictive devices in international trade. It is partly for this reason, therefore, that the Soviet Union has displayed little interest in the programs for "international economic collaboration." In the postwar period these programs are directed toward a mitigation of import restrictions; a state trading monopoly has no import restrictions except its own buying policy—which theoretically could be completely restrictive. As a monopolist seller, a state trading agency may encounter restrictions imposed by importing countries, but this is not important for the Soviet Union because its exports are almost negligible.

State trading was resorted to on a large scale by belligerents during World War II. The United States government purchased large amounts of raw materials from abroad, and with other allied governments conducted

extensive import and export transactions. As a peacetime technique of international trade control, however, state trading is largely limited to collectivist countries.

Other Controls

The types of controls described above are directed principally toward the limitation and regulation of imports. Export controls may also be involved. International commodity agreements and cartel arrangements provide for the regulation of exports. State trading comprises by definition both export and import control by government. In general, however, countries normally impose but few restrictions on exports and many even encourage exports by subsidies and other means. Since exports create balances of foreign exchange and imports reduce these balances, commercial policy tends to encourage selling and curb buying. Export controls in the form of export licenses were widely used by most countries during and directly after the war as a means of conserving and rationing scarce supplies of sugar, vegetable oils, steel, and other products among consuming nations. However, export controls offer few problems under normal circumstances; for the real burden on world trade is imposed by restrictive import policies prevailing in all trading nations.

TRENDS IN WORLD TRADE POLICY

Before the outbreak of World War I protective tariffs did exist, but since they were comparatively low and were not changed frequently, these tariffs were much less restrictive than those of the interwar period. Trade before 1914 was free to move multilaterally over tariff obstacles; foreign exchange was freely available to those who had payments to make anywhere in the world. The growth of governmental intervention during and after World War I led to an intensification of interventionist policies in foreign trade and finance. Collectivist movements were accelerated by the economic and financial disturbances of the 1930's. In the face of the general decline in trade and because of domestic unemployment each country sought to protect itself. Moreover, the military expansionist policies of Germany and Japan were tied in with foreign trade and exchange controls in order to strengthen their internal economies while weakening those of their neighbors. The effort to reduce unemployment found expression in state-planned economic programs in many countries that required control over imports. The contracting world markets led private producer interests to seek protection

through tariffs, commodity agreements, import quotas, and cartel arrangements. World War II further disrupted international trade, and in its aftermath collectivist governments increased in number and influence.

Thus, in the eyes of the old-fashioned free trader, world commerce is enmeshed in a tangle of bureaucratic red tape that lessens the productivity of economic resources, hampers trade, and retards the growth of wealth. The world, indeed, has moved far from the condition of relatively free trade that characterized the mid-nineteenth century—a condition that emerged slowly from the interventionism of the mercantilist era. Because of wars, depression, and political upheavals, the world has reverted to the practices and views of mercantilism, but with at least this difference: the present techniques of restrictive control are more efficiently organized, more effective and comprehensive.

SOME ECONOMIC CONSIDERATIONS

The economic aspects of foreign trade and commercial policy may be discussed under two headings: (1) the allocation of economic resources, and (2) the influence of foreign trade upon the level of employment. The first problem engaged the attention of economists from the close of the eighteenth century until the pre-World War II period of depression, when it was overshadowed by a preoccupation with the effects of foreign trade upon employment and national income. In attempts to revive employment or to prevent further declines, governments adopted and extended many of the trade controls that have been noted here. Foreign trade control was viewed as one of the means of mitigating depression. It is significant that the earlier interest in the international allocation of economic resources developed in a period when there were no serious doubts as to the ability of countries to maintain relatively full employment.

The Allocation of Economic Resources

From the latter part of the eighteenth century, and especially following the publication of Smith's *Wealth of Nations*, economists centered their attention on the way in which the market mechanism—the cost-price, supply-demand relationships—apportioned economic resources among the many uses for them. It was held that the free competitive market tends to allocate these resources in accordance with their most productive (i.e., least cost) employments. Following Smith's lead, David Ricardo and John Stuart Mill in the first half of the nineteenth century developed the classical

theory of international trade based upon the principle of comparative advantage. In the absence of barriers to world trade each country would tend to specialize resources in their most economic employments; therefore, maximum productivity required international free trade; departures from this policy lessened the productivity of each nation's resources.

Consider now the effects of tariffs, import quotas, exchange controls, international commodity agreements, and cartel arrangements on international specialization and the productivity of resources. In virtually all instances these controls are restrictive; they eliminate or limit imports for the benefit of domestic industries. The tariff protects some industries which otherwise might not exist at all, or might exist only on a smaller scale. The labor and capital thus employed would be more productive if they were employed in some other fields. Protection per se is an admission of waste. Import restriction, whether its purpose is admittedly protective or not, tends to encourage relatively high-cost production. The lesser productivity of resources employed in the protected industries is reflected in prices higher than those of freely imported competitive goods. In the end, therefore, consumers pay the cost of protection in curtailed incomes and consumption.

This is the free-trade indictment of restrictive trade policies. The indictment is valid in the context assumed by the free-trade analysis, namely, a condition of relatively full employment in a world free of trade restrictions. If this condition does not prevail, the force of the criticism is weakened. Some—but by no means all—of the restrictive measures described have been adopted in efforts to mitigate general unemployment. Proponents of these measures readily admit the waste of protectionism but argue that it is preferable to employ resources uneconomically than not to employ them at all.[3]

Foreign Trade, National Income, and Employment

Like all other productive activity, producing for the export market adds to national output, income, investment, and employment. By itself, production for export involves the creation of incomes in a country without the addition of a corresponding volume of goods for sale in the country. On the other hand, imports taken by themselves involve a reduction in income in a country; the incomes (or money) used to purchase imports are transferred to foreign ownership.

[3] These proponents do not include the special pleaders for protected producer groups who demand protection at all times regardless of the level of employment.

When imports and exports balance, therefore, foreign trade has no immediate stimulative or depressive influence on the level of income in a country. The amount of income created in the production of goods for export is equal to the income spent on goods imported from abroad. An excess of imports accompanied by the accumulation of idle dollar balances to the credit of foreigners tends to be depressive unless offset by forces within the economy; an excess of exports, on the other hand, tends to be stimulative to output, income, and employment.

Under some circumstances a change in foreign trade may *initiate* expansion or contraction of national income. Nations often fashion their foreign trade policy consciously to produce favorable effects on their level of economic activity. In some countries, Great Britain for example, foreign trade is so important a source of income that control over foreign trade amounts to control over the level of domestic economic activity. On the other hand, because its foreign trade is small in relation to total income, the United States can influence internal economic activity to only a relatively unimportant degree by control over trade.

The expansion and contraction of foreign trade is in large measure *dependent* upon changes in a nation's national output and income. Although, as already noted, the rise or fall in the volume of trade may initiate similar changes in national income, it is likely that in the United States, at least, the relationship works more often the other way. As national income expands United States imports and exports increase; an increase in imports naturally means an expansion of exports in countries supplying this country, and this in turn may stimulate a rise in their national income. Conversely, as the national income of the United States declines imports fall off, and foreign suppliers may suffer reductions in their economic activity. They may, therefore, adopt foreign trade policies designed to offset this adverse influence. The connections between a nation's foreign trade and its national income are by no means simple.

Some hypothetical illustrations of these relationships may make them more clear. Suppose that the imports and exports of country A are in balance, but because internal investment prospects are unsatisfactory its labor force and capital facilities are partly unemployed. To raise the level of employment it decides to reduce imports by exchange controls, or quotas, or by tariff barriers. In what way can import restrictions increase domestic employment? This can result if consumers now spend on domestically produced goods part of the income they previously spent on imported goods.[4]

[4] This may be termed a *reduction in the propensity to import*, i.e., the percentage of total income spent on imports is reduced.

If this occurs *without affecting adversely its exports,* country A would expand its production; i.e., the aggregate demand for its national output would increase in the short run by an amount equal to the decline in imports. This increase in demand for domestically produced goods would tend to induce an expansion of investment in the industries thus affected, and by means of the multiplier and acceleration effects there would ensue a rise in national output, income, and employment. Or in other words: Country A now enjoys a favorable balance of trade; its excess of exports over imports is stimulative. This excess is a component of the country's aggregate investment; its multiplier effect is stimulative to economic activity.

As matters now stand in the illustration, country A has benefited from its restriction of imports. But is this advantage likely to be held? Tracing some of the probable consequences of A's policy suggests it is only temporary. Suppose country B suffers a loss in its export trade to A because of A's restrictive import policy. The effects of this reduction in B's exports would likely be (1) a decline in income and employment in B, as production for export falls off; (2) the imposition of import controls by B, to reduce its adverse balance of payments to A; and (3) a decline in B's imports, not only because it has less A exchange and less national income, but also because it imposes import restrictions to prevent further reductions. The combined effect of a fall in B's national income and in its holdings of A exchange, together with retaliatory restrictions on its imports, would be in the course of time a reduction in A's exports.

To return again to A, consider now the effect of the decline in its exports. It was said that its import restrictions would be stimulative provided its exports were not adversely affected. But for the reasons just given its exports can scarcely be maintained. The decline in its exports means a fall in production, employment, investment, and income in the export industries, with depressive influences spreading as reduced consumer spending affects other segments of the economy.

This might wipe out the gains previously made by the restriction of imports. If its exports declined less than its imports, A would retain part of its increment of income and employment. On the other hand, if B's retaliation is drastic, the decline in A might more than offset its earlier gains.

Both A and B might reach stable and balanced import-export relationships with each other but at a considerably reduced volume of trade. Suppose both countries succeeded in establishing industries to produce goods which each had imported from the other before A's restrictive policy had

been adopted. Production of these goods (protected by tariffs or other controls) uses more resources than the production of goods each had previously exported. If these protected industries produced the same amount of goods that had been imported before, total employment in both A and B would be larger. In each country the (more productive) export industries have declined but the (less productive) protected industries have grown; production of the same amount of goods as before would therefore require a larger employed labor force. Both A and B could attain full employment if they were willing to push the mal-allocation of resources far enough without causing price increases to throttle demand.

This discussion has limited itself to changes in national income induced by an increase in import restrictions. It has shown that one country could gain at least a temporary advantage by reducing imports, but only at the expense of income and employment elsewhere. During the prewar decade restrictive policies were frequently adopted by countries determined to improve their domestic economic activity at the expense of other countries. Countries in a position to retaliate often did so (as did B in the illustration). The trade barriers created in this nationalist process partly "insulated" one economy from another, burdened foreign trade, and encouraged the less productive employment of resources.

Consider now an entirely different approach to increased national income through foreign trade policy. Country C wishes to increase trade and *reduces its import restrictions* for this purpose. The immediate effect is depressive in C because imports now exceed exports; some of its industries suffer increased competition from foreign goods. However, foreign countries now export more than before and national incomes abroad increase as a consequence. Foreign countries therefore increase their balances of C exchange; at the same time, the increase in incomes abroad stimulates a demand for C's goods in those countries. In the course of time C's export industries expand output. Whether the net effect is stimulative in C, i.e., whether the expansion of its export industries exceeds the losses suffered by other industries from foreign competition, depends upon the amount of new investment required in the export industries. If this is greater than the disinvestment in the other industries, C's national income will rise. It is probable that this would be the case in the short period. In time, however, the stimulus to the export industries in C and abroad would weaken; their investment expansions would come to an end as they build up capacity appropriate to the higher level of foreign trade. All countries would enjoy a larger volume of exports, and by the same token, larger imports.

By reducing its import restrictions, C's labor and capital are more efficiently allocated than before.

Because these stimulative effects of freer trade are recognized, the reduction of import restrictions has been advanced in several recent international councils, such as the 1947 Geneva Conference on Trade and Employment. An increase in income may be expected from this course of action. But freer trade is no panacea for all economic ills. The principal consequence of a general lowering of import restrictions at this time would probably be an increase in productivity as resources are devoted more widely to their most advantageous pursuits. Hence the net gain may be more largely in increased standards of living rather than in employment.

The stimulation of national output through foreign trade may come about in still another way, namely, by the extension of loans to foreigners. If the foreign economic policy of the United States encourages the extension of loans to foreigners by its nationals, or by the government, the immediate effect is stimulative. A loan provides foreigners with dollar exchange. If the borrower uses the funds to purchase American goods, export industries benefit and the effects are felt on many segments of the economy. Even if the foreign borrower uses the funds to pay debts to another foreigner, the second foreigner is then in possession of American money and in a position to purchase American goods. The effect of a foreign loan or investment on national income is measured by the multiplier; an investment of this type, like domestic investments, has a leverage effect, as described in Chapter 6. Many economists place considerable emphasis on foreign investments as means of absorbing or offsetting a part of the substantial volume of savings set aside annually in this economy. Those especially who fear inadequate domestic investment opportunities are prone to lay heavy emphasis on foreign loans as a means of maintaining relatively full employment. Again, the principal effect of extensive loans in the long run may be increased productivity and living standards, and not increased employment.

The repayment of foreign loans, however, has a depressive influence. As already shown, repayment of foreign loans (and the payment of interest and dividends) requires the importation of goods in large volume or other debit transactions in the foreign balance. Repayment of foreign loans is no different in its income effect from the repayment of any other type of loan. For example, a loan to build a house stimulates construction and allied trades; repayment of the loan per se creates no employment. It merely lessens the borrower's spendable funds, transfers them to the lender who may or may not spend (or invest) the money. Foreign lending therefore is stimulative only in the period during which the proceeds of the

loan are being spent by the borrowing nation. Indirectly, however, another stimulus may be felt. If the loan encourages an expansion of production and income in the borrowing nation, this increased income may induce greater purchases of American goods, thus encouraging production and income in this country. But eventually the repayment of loans must involve a reversal of favorable import-export relationships. Otherwise, the loans will become unintended gifts. Large-scale gifts or relief payments, such as those made by the United States in 1946 and 1947, are also stimulative while being spent and entail no subsequent depressive tendencies.

In the three illustrations given previously, the purpose of foreign trade policy was stimulation of domestic employment, output, and income. But foreign trade policies may also be defensive, designed to prevent adverse influences from abroad on domestic employment. For example, if the United States is a principal export market for the products of country D, the economic well-being of the latter is largely dependent upon the state of economic conditions in the United States. When the United States is prosperous, its imports from D are large, and D enjoys prosperity. Its exports to the United States provide it with ample dollar exchange to buy needed goods from this or other countries. However, when national income in the United States declines, imports from D fall off, D's export industries become depressed, and therefore D's general level of employment, income, and output. As D's exports fall, its dollar deposits in New York decline; its means of purchasing needed imports are consequently reduced. Threatened with the exhaustion of its dollar balances, D might impose exchange controls and otherwise curb its imports.[5] Many countries followed this course of action during the 1930's as United States' (and other countries') imports dropped to low levels. Under these circumstances import restriction was designed largely to conserve foreign exchange balances (and gold reserves), and to cushion the effect of depression abroad on the domestic economy. By restricting imports, D causes some shift of consumer demand from imports to domestic goods, thus encouraging an expansion (or lessening the decline) in some segments of its economy. As with B above, it attempts to gain at least partial insulation from adverse conditions abroad.

Suppose now that D embarks upon a governmentally planned program of industrial growth and diversification to assure itself a high degree of economic self-sufficiency, full employment, and protection from foreign depression influences. Under such circumstances, it might be unwilling to relax its exchange controls and other import barriers because the rise in

[5] The decline in its income would produce this effect in any case but possibly not soon enough to prevent loss of its dollar balances.

income would encourage its consumers to buy increased amounts of foreign goods. Without control, or with inadequate control over imports, its domestic prosperity might cause an unfavorable balance of payments and loss of foreign exchange.[6] Since an excess of imports over exports per se is depressive, at least in the short run, D's foreign trade position would tend to offset its domestic economic plan and at the same time deplete the foreign balances needed to purchase foreign goods essential to its high level of production. Its foreign trade controls might therefore be regarded as a means of defending its internal economy. Many countries, including Great Britain, show reluctance to abandon exchange control for reasons similar to those that prompted D to inaugurate and retain a restrictive policy on importation.

The United States, with qualified support from many other nations, has attempted in its postwar planning to bring about a reduction in world trade barriers, especially exchange controls. In the agreement establishing the International Monetary Fund, removal of many types of exchange control is mandatory. The Fund as a source of short-term loans (or exchange) removes some of the conditions that encouraged the adoption of controls. The International Bank and the proposed International Trade Organization point in the direction of freer trade. Obstacles in the path of freer trade, however, are formidable. Protected industries in this country and elsewhere can be expected to resist a lowering of tariffs. National economic planning in many nations obstructs freedom of trade and international finance, and the socialist organization of some countries actually negates it. Considerations of military security impel most nations to protect industries essential to survival in case of war. Economic and financial difficulties such as those experienced by Great Britain in 1946-47 create so great a shortage of foreign exchange that removal of exchange controls is dangerous. In spite of these obstacles some reductions in import restrictions are likely to be effected in the course of time. If the major nations of the world enter a period of sustained high levels of national income, output, and employment, the economic conditions which promoted restrictive policies in the prewar decade will be at least partly removed. Furthermore, the international financial organizations (Fund and Bank) might become partial substitutions for restrictive policies.

Of major importance, however, is the position taken by the United States, for it is at present the world's largest trader. Trends in this country indicate a long-term lessening of import restrictions. In the first place, this nation is

[6] Sweden's experience in 1946-47 is a case in point. Unparalleled prosperity led to large imports and a sharp drop in its foreign balances. This was followed by stringent import controls.

increasing its dependence on foreign supplies of raw materials. Depletion of some essential mineral resources will lead to increasing imports from abroad, especially if national output is maintained at high levels. Larger food imports may be necessary as time goes on. The industries using foreign raw materials generally favor free trade in these products. As these imports grow in importance the traditional industrial support of tariff protection is likely to abate. In the second place, industrial growth requires an extension of foreign markets for American manufactured goods and therefore increased imports.

PART VIII

PERSONAL DISTRIBUTION AND USES OF INCOME

CHAPTER 37

The Personal Distribution of Income

In earlier chapters the economic factors governing wages, interest, rent, and profits were analyzed.[1] It was shown that incomes are subject to supply and demand influences in much the same way as are other prices; indeed, this fact was described as the "pricing of the productive services." This is a theory of the *functional distribution* of income—the remuneration of the factors or resources on the basis of the economic importance of their *productive function.* Functional distribution is concerned with the market influences governing the distributive shares received by wage earners, the owners of land, the owners of capital, and entrepreneurs without regard to the *amounts* received by persons and groups.

Unlike the functional approach, the analysis of the *personal distribution* of incomes is concerned with the *amounts of income* persons and social groups receive, the inequalities of income among persons, and the causes, conditions, and consequences of existing patterns of distribution. In this analysis the distributive or functional shares—rent, wages, interest, and profits—and the influences governing them are of subordinate interest. In the next three chapters the personal distribution of income will be examined from two aspects: (1) the political and social developments which tend to modify income distribution, and (2) the economic issues involved in distribution and the developments modifying it.

[1] See Chaps. 17-21.

The distribution of income is basically governed by supply-and-demand influences—or factor productivity as measured by the market. But at the same time it is subject to "conscious social control." Much of this control is legislative, as revealed in the income-tax laws, veterans' measures, minimum wages, and social security legislation. Income distribution is affected somewhat indirectly by still other legislation, such as labor laws which strengthen (or weaken) labor unions in their bargaining with employers, and agricultural laws designed to control farm production and the prices of agricultural commodities. Thus, legislation provides income directly to certain groups: veterans, the aged, the unemployed; it reduces the incomes of other groups: those subject to the income tax; and it assures increased incomes to still other groups by controlling the supply-and-demand conditions in their favor: union workers and farmers.

Income distribution is therefore the result of the interaction of economic (market) and social controls. Incomes are also a volume of purchasing power in the hands of the public. After taxes are taken by government, the remainder (disposable income) is apportioned between spending for consumption purposes and savings. As shown before [2] the spending-saving proportions for the *economy as a whole* vary from one period to another with changes in the level of national output and income. But these ratios are governed also by the pattern of personal income distribution. That is, the proportions spent and saved by income receivers depend largely on the *amount of income* they receive; people with low incomes save nothing (as a group) while those in the high income brackets save a relatively large percentage of their incomes (as a group). Thus the pattern of personal income distribution in an economy is accompanied by markedly different spending-saving ratios by income groups; the pattern of income distribution therefore is a factor determining the level of national output and employment.

So it is that proposals and measures which would affect the pattern of income distribution inevitably bear upon the level of output and employment. Indeed such measures as old age and unemployment benefits are often urged and justified on the claim that they will bring about increased levels of output by changing the spending-saving ratios.[3]

[2] See pp. 123-25 for the discussion of this point.

[3] See the discussion in Chap. 7 summarizing the mature economy thesis which urges high taxes on large incomes, increased wages, and social security payments as means of effecting some redistribution of income to bring about a higher propensity to consume. Some who do not subscribe to the mature economy view also regard an extensive social security program, minimum-wage legislation, and similar means of affecting a wide distribution of purchasing power as necessary on economic grounds.

Not only is aggregate consumer demand, or expenditure, governed in large part by income distribution, but the make-up of this demand is also influenced by distribution. A family with $1,000 a year spends nearly all of it on food, rent, and clothing, while a family with $100,000 a year spends a relatively small percentage on these items (although more in total, of course), and a relatively large percentage on education, travel, amusements, and the so-called luxury goods and services. A distribution and expenditure of income showing such differences brings about an allocation of resources corresponding to these widely differing types of expenditure; a distribution much less unequal would bring about a different allocation of resources. Any measure, economic or social, which effects some redistribution of income therefore will change the composition of consumer demand and the allocation of resources which meets that demand.

The personal distribution of income and the social legislation that attempts some modification of it are an integral part of the great theoretical problems of economics: the level of national output, the allocation of economic resources, and the prospects of economic progress. Most of the income distributed—national income payments—is created by the process of producing national output. The disposition of this income between saving and spending varies with the size of the national output and with the pattern of income distribution; and in its turn, the saving-spending relationship affects the level of output and the prospects of long-term economic expansion.

Although these large economic issues make income distribution a vitally important subject for the student of economics, they are not the questions that absorb the attention of the ordinary man. His interest is usually limited to how much income he is getting, what he might get or thinks he should get, and what his friends and acquaintances receive. These are intensely practical interests. A person's income governs his standard of living—the necessities, comforts, and conveniences of life which he can buy for himself and his family. It determines, too, the extent to which he can save and assure a measure of comfort and security in later years. The scale on which he can spend and save, moreover, tends to establish his position in the community, for the appraisal of a person by others is sometimes largely influenced by the size of his income. Pride may be a property of the poor, but prominence seldom is.

It may be that in "getting and spending, we lay waste our powers," but for most people no socially acceptable alternative is open; necessity compels the bulk of the adult population to devote most of its waking hours to these

pursuits. This necessity stems from the basic fact of *scarcity*—of goods, services, and resources—without which there would be no distribution "problem," and in fact no economics at all. Since goods and services are scarce, people need income to buy them, and the quantity purchased is their *real income.* It is this that determines their material well-being and, to a large extent, their standing in the community.

These considerations often dominate social actions in an economy. Some men join unions to get higher wages, others seek tariff protection and monopoly advantages for larger profits, others urge governmental subsidies and higher prices for increased farm income, still others seek old-age benefits, unemployment insurance, and public works to prevent loss of income. In the ceaseless struggle for income, strikes are called and strikes are broken, political pressure is exerted and countered, advertising programs are built up and torn down; violence, negotiation, lobbying, and publicity are employed—all in order to change the existing distribution of income payments and the goods and services comprising real income. While the individual may not be interested in the larger economic issues involved in distribution, his actions to change it in his favor are the substance of the social and political forces which help decide them.

THE CLAIMS TO INCOME

The claims of individuals to most of the total income payments arise from their *current participation* in the process of production. Participation is (1) the rendering of personal services (i.e., time and effort) for the production of national output, such as the services of laborers, salaried persons, farmers, doctors, lawyers, teachers, individual business enterprisers, salesmen, and others; and (2) making assets or property available to others for purposes of receiving an income, such as money lent at interest, and land, buildings, or equipment rented. Income may therefore be derived from *personal services* and the *services of land and capital.* Earlier these sources were described as *services of the productive factors,* with the payments for, or income paid to, these productive services governed basically by supply and demand.[4] These incomes, comprising the bulk of total income received, include wages, salaries, fees, commissions, profits, interest, rent, and royalties.

[4] However, as stated above, the supply and demand factors may be influenced by political and social forces, as when union organizations supported by government are able to increase wages through this support; or when farm income is increased by governmentally sponsored reduction in farm output to increase agricultural prices.

Other incomes, small in relation to the total, are paid to nonparticipants, those who receive incomes but do not currently produce. For example, veterans' benefits and payments constitute income arising from a social right or claim as defined by law. Payments made from public funds to the aged, to dependent children, and to the disabled and the unemployed, are similarly *incomes based on social rights as defined by law.* These various payments made by government to nonparticipants are not all on the same footing, for some (old-age and unemployment benefits) are financed in part by previous taxes or contributions paid in earlier years by the persons receiving these benefits, while other payments (relief and aid to dependent children) presuppose no previous contributions by the recipients. Apart from private charity, which is a small amount of the total, incomes paid to nonparticipants have in common these characteristics: government establishes the claim to income by law and determines the amount, the period of time it is received, and the conditions on which it is paid, i.e., eligibility.

Governmental programs which establish claims to income represent a direct intervention by government in the personal distribution of income. For the most part these incomes are paid to persons, such as the aged and unemployed, who no longer receive an income from other sources, or who are considered to receive an inadequate income. The government thereby establishes a minimum or a "floor" to individual income, and in doing so mitigates the hardship resulting from loss of income.[5] In doing this, government achieves a measure of redistribution of income. In terms of real income this means that nonparticipants consume more goods and services than would otherwise be possible; and participants have less to consume by this amount. In terms of money income it means in the long run a transfer—effected primarily by taxation—from those who earn income in current production to those who do not. These effects do not necessarily mean a loss in either real or money income to the participants; a measure of redistribution may under some circumstances increase total consumption and real levels of employment and income in the economy. That is, spending-saving relationships may be changed in favor of increased consumption through the payment of such incomes; this would tend to stimulate total production if, otherwise, investment opportunities were inadequate. On the other hand, the reduction of savings when long-term investment prospects are favorable acts to reduce productive capacity and consumption in the future.

[5] These payments are part and parcel of the general trend toward *security* through governmental means; later chapters introduce this matter in considerable detail and show that the "quest for security" pervades virtually all segments of the economy and is manifested in both public policy and in the practices of private organized groups.

Redistribution and National Output

What are the limits to redistribution of income, and how does redistribution influence the national output when income from participants is transferred to nonparticipants? It is clear that those participating in current production receive incomes based on productivity and the scarcity of their services. The low wage of a common laborer and the high salary of an electrical engineer roughly measure the differences in their respective productivities. Some persons and groups enjoy high incomes because their numbers are small in relation to the demand for their services, but the supply of services is not fixed; constant striving for high incomes by newcomers to an occupation acts as a limiting factor on incomes.[6]

But the incomes of the nonparticipants are not governed directly by economic factors. Political influences and the humanitarian sentiments of the public, rather than productivity, determine these incomes. The size of the individual income, the number of persons eligible by law to claim such income, and therefore the total amount of income distributed to nonparticipants (and transferred from others) are immediately determined by the relative political importance of the recipients and the political philosophy of a government administration, as well as the social consciousness of the public. There is no direct economic influence limiting the size of the individual amounts and the total amount paid as nonparticipating income. The individual incomes may be $25 a month, or $125 a month, or $50 every Friday—or any amount the legislature decides; and this amount may be expected in general to relate to the political importance of the group or groups involved.

The matter of what determines incomes is important to the questions posed here. Participating incomes—wages, salaries, interest, dividends, and profits—are *incentives* as well as measures of productivity. Nonparticipating incomes conceivably can be *deterrents* to as well as derivations from productivity (of participants). The two claims on income are basically conflicting, no matter how justifiable socially the nonparticipating claims may be. As long as the total volume of nonparticipating income is of no consequence, the issue is unimportant. But as government provides greater amounts of such income, the economic consequences of redistribution force themselves to the fore. Redistribution under some circumstances

[6] The fact that many monopolistic practices are invoked to protect an occupation—particularly one that pays relatively well—testifies to the force of this competition; and in the long run it is probable that the monopolistic practices are no more than partially successful.

may be stimulative to the economy, but if the volume of nonparticipating incomes is large the transfer *impairs incentives to work and to invest* and depresses national output, income, and employment. And since nonparticipating incomes are immediately governed by political developments and social sentiments, there is no assurance that these adverse consequences will always be avoided. Indeed, this generalization may apply to *all* income which is either directly or indirectly influenced by political developments.

What, then, are the desirable criteria for distribution of income? Since this society places emphasis on increased (though not equal) material welfare for all persons, an ideal distribution of income would have to (1) maximize output (real income),[7] and (2) distribute this output (and therefore income) widely to avoid depressed standards for any segments of the population. This objective requires long-term economic expansion, but what assurance is there that the personal distribution of income will meet these criteria?

The economic factors which distribute income roughly in accordance with the productivity or importance of each agent of production do not by themselves—at least not in any short period of time—assure the attainment of this objective. The causes of this are:

1. The incomes distributed may not lead to a spending-saving relationship which is appropriate to maximum levels of output. Under some circumstances savings may be excessive, under others, deficient: an appropriate relationship is largely a matter of chance.
2. Because of age, disability, or inability to find employment, large groups of persons get incomes inadequate to sustain standards of well-being regarded by public sentiment as essential, when distribution is governed by economic influences alone.

Because the second of these causes has been forcibly impressed on the political consciousness of the nation, social security and related programs have been instituted to alter the distribution of income. Is it certain, however, that the distribution of income, as modified by nonparticipating claims to income, will assure the social objective of long-term growth in material well-being? As noted, this objective combines maximum output (or income) and a wide distribution of this output (or income). The nonparticipating incomes achieve a wider distribution of output and income than would be found without them.[8] But this fact does not, in itself, assure the attainment of

[7] Given the resources at hand and preferences for commodities, services, and leisure.
[8] This is also one of the purposes behind minimum-wage legislation, farm legislation, and labor legislation.

maximum output (or income) in the long run. The economy requires an abundance of investment opportunities and adequate incentives to invest in order to attain maximum output over the long period. Since the satisfaction of consumer demand is the chief (indirect) objective of investment, the economy must maintain a nice balance between the income used for consumption and the funds set aside for investment. The political and social influences behind the redistribution of income may at times be too weak to bring about an increase in consumption necessary to attain this balance; i.e., they may not be able to redistribute income sufficiently to bring about the spending-savings relationship currently required for maximum levels of national output, income, and employment.

At other times, these political and social influences may be too strong; redistribution is carried to the point where it impairs incentives, particularly incentives to invest. This impairment is possible if (1) the amount of public money needed for the payment of nonparticipating incomes is large, thus requiring large volumes of tax receipts, and (2) as a result of the "ability-to-pay" principle high incomes are taxed heavily. This latter point is important because the high-income group derives a large part of its income from investment, accumulates a large portion of the individual savings in the economy, and is an important investor group. High income taxes, as seen earlier, reduce prospective returns on investment and therefore affect expectations adversely. Even if incentives are not impaired, too little may be left of these high incomes for savings, i.e., too little in relation to the needs of investors for savings.

In short, the political and social forces by themselves do not assure a redistribution (if this in fact is needed) appropriate to maximum levels of national output, income, and employment. Redistribution may fall short of, or exceed, the requirements of the economy at any given time. And, as stated above, the economic influences by themselves may not assure a distribution of income appropriate to maximum levels of output, income, and employment. This is so because the political and social influences on the one hand, and the economic influences on the other, tend to emphasize different aspects of the social objective. The economic influence, by apportioning income in terms of *productivity,* emphasizes the unhampered growth of output and income. The political and social influences effect a redistribution in terms of individual and group *needs,* thus emphasizing the wide distribution of income and output.

The contrasts cannot be drawn sharply but they do bring out an underlying conflict in methods to attain in an economy the social objective:

maximum output and wide distribution of output.[9] In the early growth of industrial capitalism the prevailing economic philosophy of mercantilism favored a rapid growth of production or national output. Consumption was held down by the payment of very low wages to the "lower orders," while those who enjoyed high incomes were exhorted to save and invest. In this period a wide distribution of income would have retarded economic expansion. Similarly in Russia—and to a lesser extent in prewar Germany—governmental policy of industrial expansion keeps consumer income and consumption at very low levels; the consumer goods industries are retarded while capital goods industries are built up. The destruction of property wrought by war imposes presently the same choice—the rate of reconstruction depends upon the extent to which current standards of well-being (i.e., the propensity to consume) are held down in the countries affected.[10]

On the other hand, a well-developed industrial economy such as the United States is not faced with so sharply defined a contrast between current well-being and long-term economic expansion. Productive capacity here is such that high (in relation to other nations) and increasing standards of well-being are consistent with economic expansion. Indeed, since private investment in this economy is made in anticipation of increased consumer demand, rising standards of well-being are essential for the expansion of investment and output in the long run. Income distribution plays an important part in this expansion; it must be so constituted as to encourage an adequate volume of current consumption and allow sufficient savings to be set aside to finance investment and production for future increases in consumption.

THE DISTRIBUTION OF INCOME

Following this brief discussion of the large economic issues underlying distribution the discussion now turns to a review of some of the statistical information on the subject. Figures are available showing the *total amount* of income paid out during the last several decades, but figures on the *distribution of the annual totals* among the millions of income receivers exist for only a few years. The years reviewed in the following pages are 1929 (peacetime prosperity), 1935-36 (depression), and 1943 (wartime prosperity). For these years information is available on the following:

[9] Again, maximum output is interpreted in terms of preferences for commodities, services, and leisure.

[10] If a country borrows from abroad and imports capital facilities its expansion may be furthered by such imports, and it may be able to devote more of its resources to the production of consumer goods.

1. *Distribution of income,* as revealed by two measures:
   a. The number of families and single persons found in each income group or class; for convenience, income classes are arranged on the following basis: 0-$1,000, $1,000-$2,000, $2,000-$3,000, etc., with the highest group for present purposes receiving $10,000 a year and over. This is *distribution of income receivers by income classes.*
   b. The percentage of total income received by these income classes, or by certain percentages of the income receivers. In 1943, for example, all income receivers are divided into tenths, ranging from the highest 10 per cent to the lowest 10 per cent, and the percentage of total income going to each tenth is shown. This is the *proportionate distribution of income.* It provides a measure of the extent of income inequality in the economy.
2. *Income use*—how incomes are saved, spent, and paid out in taxes. The income studies reveal vastly different saving-spending ratios for the various income classes; and this information brings out the relationship between the personal distribution of income and the saving-spending ratios applicable to national income. Both class and national ratios are related to the level of national output, income, and employment. (This is discussed in Chapter 38.)
3. *Major types of consumer expenditure*—for food, clothing, shelter, etc.—are shown by income classes. Among these classes there is considerable difference in the character or composition of consumer expenditures. This part of the income studies is closely related to the allocation of economic resources. (This is discussed in Chapter 38.)

Distribution of Income in 1929 [11]

Table 77 shows the distribution of the 1929 national income of almost 93 billion dollars among the 36 million income receivers of that year. These comprised over 27 million families of two or more persons, together with nearly 9 million single persons. The average income of this combined group was $2,550; the average family income was $2,800, the average individual income was $1,760. Half of the nation's families in 1929 received less than $1,700.

The distribution of income receivers by income classes is presented in

[11] Income estimates, as well as estimates of income disposition in 1929, are not comparable to the estimates shown in Chaps. 4-6. The differences in aggregates are not important for the purposes of this and the succeeding three chapters.

Table 77; column 1 shows the selected income classes; column 2, the number of income receivers in each class; column 3, the percentage distribution of income receivers. The predominance of the two lowest income groups and the sharp reduction in numbers as income increases are noteworthy. Sixty-five per cent received incomes of $2,000 and less; only 2.1 per cent received $10,000 and more. In terms of aggregate income, the 2.1 per cent received as much as the 65 per cent.

Table 77 ESTIMATED INCOME DISTRIBUTION IN 1929

	Income receivers		Income		
1. Income class in dollars	2. Thousands of persons	3. Per cent	4. Millions of dollars	5. Per cent	6. Average income in dollars
0–1,000	10,020	27.5	5,109	5.5	510
1,000–2,000	13,775	37.8	20,092	21.6	1,460
2,000–3,000	6,059	16.6	14,637	15.7	2,430
3,000–4,000	2,695	7.3	9,261	10.0	3,420
4,000–5,000	1,350	3.7	6,005	6.5	4,440
5,000–10,000	1,816	5.0	12,123	13.0	6,780
Over 10,000	747	2.1	25,723	27.7	34,500
All classes	**36,462**	**100.0**	**92,950**	**100.0**	**2,550**

Source: M. Leven, H. G. Moulton, and C. Warburton, *America's Capacity to Consume*, Brookings Institution, Washington, D. C., 1934, pp. 227-30.

The fact that 10 million families and single persons received an average of $510 during the prosperous year of 1929, and that 13.7 million averaged under $1,500, does not accord with many of the popular notions of prevailing incomes under conditions of relatively full employment. Nor does it accord with many of the estimates of "adequate" income current at the time. The question of income adequacy always arises in a discussion of the personal distribution of income.

A number of estimates have been made of "minimum health and decency incomes" and "minimum standards of well-being" and other variations of the adequacy concept. The Brookings Institution[12] estimated $2,500 as the minimum to provide a family with "moderate well-being" in 1929. Its estimates show 19.5 million families, or nearly three quarters of all families, with incomes of $2,500 and less. However, one cannot conclude from this estimate alone that all or most of these 19.5 million families were below the level of moderate well-being. A diversity of conditions in this country makes such generalizations somewhat dubious.

[12] This is a privately endowed research organization in Washington, D. C., well-known for its studies of economic problems.

For example, a family income of $2,500 has not the same purchasing power in all parts of the country. In rural areas living costs are generally lower than in cities, and lower in the milder climate of the South—both in the cities and the country—than in the North. Moreover and very obviously, a family of two can live much better than a family of six on $2,500 a year. Location, climatic factors, regional differences in modes of living, and differences in family size are some variables which must be applied to incomes to determine adequacy.

These qualifications are borne out in part by the estimates under discussion. Farm family income averaged $1,240 in 1929, town and urban families, $3,226.[13] Ninety per cent of farm families earned $2,500 or less, compared with 71 per cent for the nonfarm families. And relatively more of both groups earning less than $2,500 were in the South than in the North. Thus it is reasonable to suppose that families earning less than $2,500 were concentrated in the lower-cost-of-living areas, those earning more than $2,500 in the higher-cost-of-living areas—with the concentration in both cases measured as a percentage of families.

The proportionate distribution of income in 1929 is shown in column 5 of Table 77. Twenty-seven per cent of total income went to the 2.1 per cent in the $10,000 and over group; and 27 per cent of income went to 65 per cent of income receivers in the $2,000 and under groups. The top 10 per cent of the income scale received almost as much of the total income as the remaining 90 per cent. More striking still in portraying the extent of income inequality in 1929, is the fact that the 36,000 families receiving $75,000 or over, who comprised only 0.01 per cent of those getting incomes, received as much as the 42 per cent on the lower end of the income scale. Stated differently, these 36,000 families received as much as the 15 million families on the lower end of the scale.[14]

The distribution of income receivers and the concentration of income revealed in Table 77 applied at a time when national income was near its maximum under peacetime conditions and with the economic resources at hand.[15] This, it will be recalled, is one meaning of the term relatively full employment.

[13] The farm income estimates included an allowance for rent of the farm home and for farm produce consumed, as well as earnings of farmers from nonagricultural work.

[14] For a fuller account of income distribution in 1929 see M. Leven, H. G. Moulton, and C. Warburton, *America's Capacity to Consume*, Brookings Institution, Washington, D. C., 1934, Chap. 5.

[15] "Maximum" income in this sense means "real" income.

Income Distribution in 1935-36 [16]

Table 78 shows the distribution of income in a period of depression for the same income classes given in Table 77. National income in this year was 59 billion dollars, well below the 93 billions of 1929; and the average income was $1,500, in contrast to $2,550 for 1929. Part of this decline was occasioned by the lower level of prices in the depression period.

Table 78 ESTIMATED INCOME DISTRIBUTION IN 1935–36

	Income receivers		Income		
1. Income class in dollars	*2. Thousands of persons*	*3. Per cent*	*4. Millions of dollars*	*5. Per cent*	*6. Average income in dollars*
0–1,000	18,359	46.5	10,806	18.2	590
1,000–2,000	13,920	35.8	19,573	33.0	1,400
2,000–3,000	4,434	11.2	10,576	17.9	2,390
3,000–4,000	1,354	3.5	4,599	7.7	3,400
4,000–5,000	464	1.1	2,044	3.5	4,430
5,000–10,000	596	1.5	4,092	6.9	6,860
Over 10,000	328	0.8	7,566	12.8	23,000
All classes	**39,458**	**100.0**	**59,258**	**100.0**	**1,500**

Source: *Consumer Incomes in the United States,* National Resources Committee, Washington, D. C., 1938.

The general effects of depression on all income groups are clearly revealed in these figures. When the distribution of income receivers by income classes in this year is compared with that in Table 77, it is evident that a substantial downward shift occurred over the period. Over 18 million income receivers—8 million more than 1929—representing 46 per cent of the total, were crowded into the lowest income class. All groups above $3,000 had shrunk considerably, with the largest percentage reductions occurring progressively higher on the income scale.

The shift downward in incomes is one of the manifestations of depression. Unemployment and underemployment are widespread, forcing large numbers of workers into the lowest income class. After 1930 those who held their jobs often worked at lower wages; many wage and salary workers had to take other lower paid employment; farm incomes dropped with farm prices; the profits of enterprise were reduced and in many cases wiped out; divi-

[16] Like the 1929 estimate, the 1935-36 estimate differs from the aggregate shown in Chaps. 4-6. The difference does not affect the usefulness of the estimates for present purposes. Also some slight differences between the methods of estimating 1929 and 1935-36 incomes and uses should be noted here.

dends and rent payments declined—few incomes went through this period unscathed.

The 1935-36 income study reveals a relatively heavy concentration of certain population groups in the lowest income class. (1) More than 90 per cent of the Negroes in Southern rural areas and towns and over 80 per cent in all other sections of the nation received $1,000 or less. (2) Sixty per cent of the single persons received $1,000 or less. (3) Over half of the farm families were in this income class.[17] These three groups accounted for probably two thirds of the total population in the lowest income bracket. This illustrates the fact that in depression as well as in prosperity the incidence of low income is uneven—and persistent—in terms of regional, occupational, and socioracial groups in the economy.

Next arises the question: What changes, if any, occurred in the *proportionate distribution* of income during the prewar depression? Was income distribution more, or less, unequal in this period? In general, there was a reduction in the degree of inequality as shown in Table 79.

Table 79 COMPARISON OF INCOME, SELECTED GROUPS

Income receivers	Per cent of total income received 1929	1935–36
Upper 5 per cent	36	27
Upper 10 per cent	47	36
Upper 15 per cent	51	49

The upper 5 per cent of the income receivers in 1935-36 received $3,400 or more; in 1929, $6,000 or more. This again illustrates the change in the distribution of income receivers during the depression years. The lessening of inequality during the depression period brought scant comfort to those who decry the extent of income inequality in this country, for it meant merely that while the poor became poorer, they suffered a relatively smaller decline in income than did those on the higher reaches of the income scale.

Although less than in 1929, income inequality remained pronounced throughout the depression period. The highest 5 per cent of the income receivers received 27 per cent of total income—almost as much as the lowest 60 per cent of all families and persons. Ten per cent of the 1935-36 income went to a group comprising only 0.5 per cent of the income earners, a total of only 197,000 families; whereas an equal amount of income supported 3.8 million families and individuals (about 9 per cent of the total) on the lower end of the scale.

[17] The low-income Negro farm families partly account for this fact.

Table 80, showing the proportionate distribution of income, brings out the fact of inequality in somewhat more precise terms than in Table 79. The 39,450,000 income receivers are divided into tenths, ranging from the highest tenth to the lowest tenth. Column 4 shows the amount of income going to each group of 3,945,000 families and single persons; in column 5 these amounts are shown as percentages of total income. The 3,945,000 receiving the highest incomes obtained over 21 billion dollars, or 36 per cent of total income for the year; this was nearly as much as received by the seven groups (70 per cent) at the bottom of the scale which were comprised of nearly 27 million families and single persons.

Table 80 PROPORTIONATE DISTRIBUTION BY TENTHS: ALL INCOME RECEIVERS 1935–36

	Income receivers		*Aggregate income*		
1. Proportion of families and single persons	*2. Number of families and single persons in thousands*	*3. Income range in dollars*	*4. Amount in millions of dollars*	*5. Per cent in each group*	*6. Cumulative per cent*
Highest tenth	3,945	2,600 and over	21,452	36.2	100.0
9th	3,945	1,925–2,600	8,593	14.5	63.8
8th	3,945	1,540–1,925	6,815	11.5	49.3
7th	3,945	1,275–1,540	5,511	9.3	37.8
6th	3,945	1,070–1,275	4,444	7.5	28.5
5th	3,945	880–1,070	3,911	6.6	21.0
4th	3,945	720– 880	3,259	5.5	14.4
3rd	3,945	545– 720	2,548	4.3	8.9
2nd	3,945	340– 545	1,719	2.9	4.6
Lowest	3,945	Under 340	1,007	1.7	1.7
Total	**39,450**		**59,259**	**100.0**	

Source: Consumer Incomes in the United States, op. cit.

Under conditions of relatively full employment and high levels of income an appreciable change in the amount of income received by one segment—those below $1,000 a year, or above $10,000, or any other group—would tend to be at the expense of other groups. This would have been the case in 1929, for total income was near maximum levels. However, when an economy is operating considerably below full-employment levels, as in 1935-36, an increase in the income of any one group need not involve a reduction of incomes elsewhere in the economy. In fact, an increase in the average income of one group usually signifies an increase in all other groups. Under such circumstances the economic welfare of an economy

depends less, if at all, on redistribution than on an increase in total income.

In the years following 1935-36, particularly during the war years, national output and income reached higher levels. The income distribution estimates for 1943 clearly show the effects of wartime prosperity, both on the distribution of income receivers by income groups and on the proportionate shares of income.

Income Distribution in 1943

Total national income in 1943 was approximately 117 billion dollars, almost double the 1935-36 figure. The average income per recipient was correspondingly higher: $2,910 compared with $1,500. Income data are shown in Table 81 on the same basis as in Tables 77 and 78.[18] The far-reaching influence of the war prosperity on individual incomes or the distribution of income receivers is brought out clearly by a comparison of column 3 of Table 81 with the same column in Table 78. Thus in 1943, 20 per cent of income recipients received under $1,000, compared with 47 per cent in 1935-36; for the $1,000-$2,000 group the percentages were 27 and 36, respectively. All the groups from $3,000 up increased manyfold; nearly four times as many people were in the $3,000-$4,000 group, seven times as many in the $4,000-$5,000 group, etc. An upward shift along the income scale occurred during the war years, induced by increased employ-

Table 81 ESTIMATED INCOME DISTRIBUTION IN 1943

	Income receivers		Income		
1. Income class in dollars	2. Thousands of persons	3. Per cent	4. Millions of dollars	5. Per cent	6. Average income in dollars
0–1,000	8,100	20.0	5,026	4.3	620
1,000–2,000	11,043	27.4	16,509	14.1	1,500
2,000–3,000	8,253	20.5	20,241	17.2	2,460
3,000–4,000	4,925	12.2	16,994	14.5	3,450
4,000–5,000	3,265	8.1	14,446	12.3	4,430
5,000–10,000	3,800	9.4	24,703	21.0	6,500
Over 10,000	964	2.4	19,481	16.6	20,209
All classes	**40,235**	**100.0**	**117,400**	**100.0**	**2,910**

Source: Civilian Spending and Saving, Office of Price Administration, Washington, D. C., 1943.

[18] These yearly figures are not wholly comparable, but for present purposes complete comparability is not essential. It should also be stated that part of the increase in 1943 over the earlier period is accounted for by price increases. Again the total for this year differs from more recent estimates shown in Chaps. 4-6.

ment, more continuous and longer work, higher wages and salaries, better farm prices, and enormously increased returns from trade and industry.

In comparison with 1929, the total national income was about 20 per cent greater in 1943. In this year the economy was entering a period characterized by an extended use of economic resources, and consequently income was high.

Although the wartime expansion greatly changed the incomes of most of the population—as is evident in the different distribution of income receivers compared with the other years—only a small change occurred in the proportionate distribution of income. This is a striking fact. The doubling (over 1935-36) of the flow of income to the 40 million income receivers might be expected to affect appreciably the proportions going to these people. But in Table 82, showing the distribution by tenths for 1943, the percentage of income going to each 10 per cent of the income receivers is very similar to that in Table 80 for 1935-36. The highest tenth in 1943 received 34 per cent of total income, compared with 36 per cent in 1935-36; the lowest tenth received 1.4 and 1.7, respectively. Income inequality in the highly prosperous period of 1943 was therefore nearly as pronounced as in the depression period; the great increase in income was distributed in such a way that each group remained in about the same relative position as before. The highest ten per cent of the income receivers obtained over 40 billion

Table 82 PROPORTIONATE DISTRIBUTION BY TENTHS: ALL INCOME RECEIVERS 1943

Income receivers			Aggregate income		
1. Proportion of families and single persons	2. Number of families and single persons in thousands	3. Income range in dollars	4. Amount in millions of dollars	5. Per cent received by each group	6. Cumulative per cent
Highest tenth	4,023	5,335 and over	40,366	34.4	100.0
9th	4,023	3,991–5,335	18,425	15.7	65.6
8th	4,023	3,138–3,991	14,234	12.1	49.9
7th	4,023	2,550–3,138	11,382	9.7	37.8
6th	4,023	2,102–2,550	9,345	8.0	28.1
5th	4,023	1,720–2,102	7,690	6.5	20.1
4th	4,023	1,358–1,720	6,207	5.3	13.6
3rd	4,023	993–1,358	4,756	4.1	8.3
2nd	4,023	634– 993	3,336	2.8	4.2
Lowest	4,023	Under 634	1,659	1.4	1.4
Total	**40,235**		**117,400**	**100.0**	

Source: Civilian Spending and Saving, op. cit.

dollars in 1943, an amount equal to that received by 65 per cent on the lower end of the income scale, and the top 20 per cent received as much as the remaining 80 per cent—proportions which were about the same as those in 1935-36.

Some Conclusions on Income Distribution

The three estimates summarized here cover widely different circumstances: (1) the peacetime prosperity of 1929, when national income was at a high level and production was carried out under normal conditions; (2) the depression period of 1935-36, when national income was low and production and employment were well below normal prosperity levels; and (3) wartime prosperity, when income, production, and employment were abnormally high when judged by previous standards.

It is readily evident from the statistical evidence that when total national income available for payment rises and falls by large amounts, changes in total income are distributed throughout the entire income range; they affect the high, medium, and low incomes in the same way, but not to the same degree. Thus when total income increases, all income groups share in the added income; they share likewise in loss when income decreases.[19] This means that the well-being of income receivers in general is governed by the level of national output, income, and employment. The larger the volume of real income, or goods and services, the larger is the total of income for payment throughout the range of income receivers.

Although the personal distribution of income, as seen in the distribution of income receivers among the various income classes, is subject to important changes, *the extent of income inequality—proportionate shares—does not change materially,* as indicated by the evidence at hand. The richest 10 per cent, for instance, takes about the same proportion in good years as in bad; the poorest 10 per cent also receive a consistently proportionate share. An estimate for 1910, for example, shows the richest 2 per cent of the population received 20 per cent of all income; the poorest 65 per cent received 39 per cent.[20] In 1943, when total money income was about four times greater, these proportions were nearly identical. About half of all income normally goes to the 15 per cent of receivers on the top of the income scale, and the other half is shared by the remaining 85 per cent.

[19] There are of course many income receivers whose incomes change little or not at all during the great swings in total income, but the extent of the shift up and down the scale is evident in the tables given in this chapter.

[20] C. R. Daugherty, *Labor Problems in American Industry,* Houghton Mifflin, Boston, 1944.

Inequality in income distribution is, it should be noted, no more pronounced in the United States than elsewhere. It may conceivably be less pronounced, although national comparisons are difficult to make. Income data for England and for some of the areas of continental Europe reveal a high degree of inequality persisting over many decades. Such income statistics as are available for Soviet Russia show an inequality for that socialist country as great as found in capitalist countries; the top 10 per cent of income receivers—government officials, plant managers, and others in the upper income class—receive about 50 per cent of the total, leaving to the 90 per cent (chiefly the farm group and wage earners) the remaining half.

Since income inequality has persisted presumably without notable change for a long period of time, the improvement in the well-being of the population as a whole in this country and elsewhere has been occasioned mainly by the great expansion in national output and income. The *total* has increased, with the proportionate distribution of income changing but little.

CHAPTER 38

The Disposition and Use of Income

The two major economic aspects of personal-income distribution are (1) the relationship of distribution to the saving-spending ratio for the economy as a whole, and (2) the relationship of distribution to the types of consumer expenditure. The first bears upon the *levels* of resource use, the second upon the *types* of resource use. As has been shown, the level of national output, income, and employment depends upon aggregate demand —of consumers, investors, and government—and the composition of this aggregate demand is in large measure determined by the spending and savings habits of income receivers. Spending and saving ratios vary greatly by income groups; therefore personal-income distribution bears directly upon aggregate demand and the level of national output. Its influence on the types of output produced and therefore resource allocation is also direct; resources are allocated so as to satisfy existing and anticipated aggregate demand, the bulk of which is normally consumer demand. The character of

this demand—or the types of consumer expenditure—varies by income groups; therefore personal-income distribution plays a part in determining the composition of national output and the types of resource use. Thus, the following analysis completes the theoretical outline presented in this book, for it describes how consumer demand for goods and services—the mainspring of all economic processes—is related to personal-income distribution and to the level and types of output.

Income paid out is both the reward for participation in the productive process (except nonparticipating income) and the incentive to production. Output thus produced by private enterprise, principally in expectation of profitable sale in the months or years ahead, comprises the bulk of total output in this economic system; the balance of national output is comprised mainly of governmental services. If national income payments aggregate, say, 167 billion dollars in a year, these comprise the principal cost of production. It is obvious that all national income payments received by the public at large have to flow back to the sources of production if this part of total cost is to be covered. When this flowback of money income is *less* than total income payments, the value of national output will decline because some producers will be unable to cover costs.[1] When the flowback is more than current income payments (because individuals redeem war bonds, draw on savings accounts, etc., or businessmen borrow from their banks or government borrows and spends) national output tends to rise, and income, employment, consumption, and savings are increased.

Since the flowback of income payments plays a large part in governing the level of output, it is essential to examine the disposition or use of income by those receiving it. How is this income used? What influence has income distribution on its use or disposition?

The answer to the first question has been given in some detail earlier. Most of total income payments are spent currently for consumption purposes, because the propensity to consume for most income receivers is high. Another part of income is paid in the form of direct (chiefly income) taxes; these funds usually are promptly spent by government in carrying out its many public responsibilities.[2] The rest of the national income is saved by income receivers. For the most part savings are deposited in banks, or put

[1] Real output may be maintained, however, if producers are willing to continue operations at lower prices and profits, or at moderate losses. If losses are heavy and widespread and threaten to continue, real output will fall.

[2] If taxes exceed government outlays the Treasury surplus will be used chiefly to retire debt. Under some circumstances this may result in a reduction in the supply of money and in incomes.

into insurance, or used to pay debts or to buy securities. Investment absorbs savings and "puts them to work," thus creating additional income (and savings). The propensity to save, or the saving ratio, is the measure of how much will normally be saved out of total income, just as the spending ratio is the measure of how much will be spent for consumption. The return of savings into the channels of industry and trade by investment is determined by long-term profit expectations. Expectations, in turn, are influenced by current profit yields (i.e., price-cost-quantity relationships), which are largely governed by current consumer expenditures (or business receipts) and production costs. This is not a horizontal process but a circular one, for not only is investment largely dependent upon current consumer expenditure, but also a part of the income making such expenditure possible results from the production of capital facilities or investment goods. This complex interrelationship between spending, saving, and production has been described earlier,[3] so it can be emphasized here only that the saving-spending ratios—the amounts saved and spent at given levels of income—are to be considered in the light of their bearing on long-term profit expectations, therefore on investment, total output, income, and employment.

Granting the importance of saving-spending ratios for the economy as a whole, the question arises: What influence does the distribution of income have on these ratios? It is clear that it has a very direct influence. Amounts of saving and spending are determined chiefly by the amount of income received. Low-income receivers as a group save nothing; their incomes are spent entirely for current consumption. Higher on the income scale savings appear; for greater incomes the percentage saved increases and accordingly the percentage spent for consumption decreases. Whether *aggregate* savings are large or small depends upon the percentage distribution of income receivers; i.e., a distribution marked by great inequality promotes large savings while a lessening of inequality may be expected to restrain savings and promote consumption. This point will be illustrated later.

Income distribution, too, affects the composition or types of goods and services purchased by consumers. The rich and the poor spend not only substantially different amounts but also buy substantially different things. An increase or a lessening of inequality therefore changes the composition of consumer spending as well as the total amount spent. Since consumer demand or expenditure largely guides the allocation of economic resources, any marked change in the personal distribution of income alters the pattern of allocation because it changes the composition of consumer demand.

[3] Chap. 5.

SAVING AND SPENDING PATTERNS

Saving and Spending in 1929

In this year 27.4 million families in the country received a total of 77 billion dollars income.[4] Table 83 shows the disposition or use of this income.

Table 83 **DISPOSITION OF FAMILY INCOME, 1929**

	Billions of dollars	*Per cent of total*
Consumer expenditures	60.7	78.3
Individual savings	15.1	20.0
Income taxes (and gifts)	1.3	1.7
Total	**77.1**	**100.0**

Source: America's Capacity to Consume, op. cit.

For purposes of analysis, the families may be divided according to income into ten groups of 2,740,000 each. Several of these groups are shown in Table 84 to bring out the diverse saving-spending ratios by different income groups.

Table 84 **DISPOSITION OF FAMILY INCOME, BY INCOME GROUPS, 1929**

Family groups		*Per cent of income*		
		1. Spent	*2. Saved*	*3. Personal taxes and gifts*
Highest tenth	$4,600 and over	60	37	3
Ninth tenth	3,100–4,600	83	17	*
Eighth tenth	2,450–3,100	87	13	—
Fifth tenth	1,450–1,700	95	5	—
Third tenth	950–1,250	100	0	—

Source: America's Capacity to Consume, op. cit.

* Less than 0.1 per cent.

The saving ratio is high for the highest tenth on the income scale; it decreases as income is less. Savings are zero for the third tenth and negative for the two lowest groups (not shown). On the other hand, the spending ratio is less for the highest tenth and increases for each succeeding group; indeed the two lowest groups consume more than their incomes—by using up savings, selling personal belongings, and going into debt. Savings become positive above the $1,200 income level; consumption equals income at this level.

[4] The estimates of saving and spending in 1929 cover *families* only.

The bearing of income distribution on saving and consumption may be brought out by assuming modifications to occur in distribution, using the ratios prevailing in 1929. Suppose that 1 billion dollars are transferred from the top income group and distributed among the three lowest groups, i.e., those receiving $1,250 and less per year. What effect would such a transfer have on savings and spending? In the first place, savings would decline roughly $370,000,000, because 37 per cent of the billion dollars would have been saved had it remained in the hands of the highest income group. Transferred to the lowest groups virtually none of it would be saved. The top group would spend $600,000,000 less for consumption, but the lowest group would spend approximately 1 billion dollars more. Thus, such a transfer would tend to reduce savings and increase consumption spending in the economy as a whole.

The illustration might be carried to the extreme. Suppose that income inequality were entirely wiped out. This would have given each family $2,800 in 1929. Families at this income level in this year actually saved 13 per cent of their income and spent 87 per cent. If the saving-spending ratio for this level of income was maintained now by all families, only 13 per cent would be saved instead of 20 per cent of total income formerly saved. Instead of 15 billion in total savings, there would only be 10 billion dollars. By the same token, 87 per cent of all income would go for current consumption purposes, instead of 78 per cent. A change in income distribution in so drastic a form would have altered materially the saving-spending ratios of almost all the population.

Next, suppose that inequality were intensified by a shift of 1 billion dollars from the 90 per cent who received less than $4,600 to the 10 per cent who received more than $4,600. In the hands of the larger group only 5 per cent of this billion dollars would be saved, and 95 per cent spent. But transferred to the highest ten per cent of the income receivers of the income scale, only 60 per cent would have been spent and 37 per cent would have been saved. Thus such a shift would have decreased consumption by about 350 million dollars, and savings would have increased by nearly 350 million dollars.[5]

The data given here and these hypothetical illustrations establish the presumption that income distribution is a governing factor on the saving-

[5] The changes in saving and spending assumed in this illustration are not exact, but only indicative of the size of the shifts in saving and spending. The transfers of the 1 billion dollars assumed here would be associated with somewhat different saving-spending ratios. That is, the addition of 1 billion to the high-income group might increase their saving ratio, and a decrease of 1 billion might lessen it somewhat. For present purposes such refinement of analysis is not necessary.

spending ratios for the economy as a whole. Income inequality promotes savings. The upper 10 per cent saved 37 per cent of their income because they had enough income to permit both large savings and high levels of consumption. The middle 10 per cent saved little and the lowest 10 per cent saved nothing at all, because they had little or nothing to spare after satisfying their current consumer needs (and possibly not much will to save). A lessening of inequality would tend in the short run to raise the consumption of the lower income groups by more than the reduction in consumption among the high-income group; alternatively, savings among the top group would be decreased by more than the increase in savings among the low-income groups. In terms of output, consumer goods production would increase, and investment goods fall; in the long run the lower level of investment (and savings) would slow up the growth of consumption. Indeed, if saving were reduced far enough, future consumption would fall.

The connection between income distribution and the saving-spending ratios for the economy as a whole has been used by some persons as a justification for income redistribution, especially those who ascribe to the mature economy thesis. As shown earlier,[6] this thesis asserts there is a tendency toward oversaving in the economy—specifically, at high levels of income the propensity to save would exceed the propensity to invest with a consequent tendency toward semistagnation in the economy. Oversavings and therefore underconsumption can only be cured, according to this view, by a reduction in the savings ratio and an increase in the consumption ratio. The means of achieving this change would include: high taxes on income that normally would be saved, high wages, a comprehensive social security system, and low prices on consumer goods to assure maximum consumption by the public at large. Against this depression-born view is set that of the expansive economy proponents, who assert that savings would be fully invested if deterrents to investment were removed, and that measures to reduce savings merely impose on the future reduced consumption standards.

Saving and Spending in 1935-36

The disposition of the 59 billion dollars of national income received by both families and single persons in this year follows patterns similar to those shown for the year 1929. The total being lower than 1929, savings and consumption were also smaller. Table 85 shows the 1935-36 use of income.

[6] See Chap. 7, pp. 153-58.

Table 85 DISPOSITION OF TOTAL INCOME, 1935–36

	Billions of dollars	Per cent of total
Consumer expenditure	50.2	84.7
Saving	6.0	10.1
Income taxes	0.9	1.5
Gifts	2.2	3.7
Total	**59.3**	**100.0**

Source: Consumer Expenditures in the United States, National Resources Board, Washington, D. C., 1939.

In 1935-36, as in 1929, the percentage of income saved was high among those on the top of the income scale and decreased successively with lower incomes. Again, the spending ratio was relatively low for the high incomes and increased successively with lower incomes. These relationships are again given for several income levels.

Table 86 DISPOSITION OF INCOME, BY INCOME GROUPS, 1935–36

	Per cent of income		
	Spent	Saved	Personal taxes and gifts
$20,000 and over	35	51	14
5,000–10,000	64	30	6
2,000– 2,500	87	9	4
1,000– 1,200	100	—	—

Source: Consumer Expenditures in the United States, op. cit.

The 6 billion dollars of savings in this year constituted a smaller percentage of total income (10 per cent) than did the savings of 1929. The downward shift of income receivers on the income scale accounts for this; there were fewer families earning enough to permit substantial savings, and there were more who did not save and who dis-saved in this year compared with 1929. However, when spending-saving ratios by income levels are compared for the two years (1929 and 1935-36) there is a striking similarity. In both periods savings were negative up to the $1,200 income; positive savings appeared above this at about *the same ratio for each income group.*

Wartime Changes

The great expansion in national income during the war years was accompanied by a marked upward shift of income receivers on the income scale. This is evident from Table 81 in the preceding chapter, where it was noted, however, that the increased income (at least for 1943) occasioned no significant change in the proportionate distribution of income (as compared with 1935-36).

As total income increased, the spending ratio declined and the savings ratio increased. Thus in 1943-44, 22 per cent of total income was saved and 66 per cent spent for consumption purposes; the remaining 12 per cent comprised personal taxes. The reduction in the spending ratio and the rise in the savings ratio may be explained by (1) the upward movement of income receivers on the income scale and (2) wartime conditions and policies of government. From the data presented in Tables 83 and 85 for 1929 and 1935-36, it is clear that a rise in individual incomes is accompanied by an increased percentage of income saved and a reduction in the propensity to consume. Apart from any other influences, then, it may be expected that the upward movement of income receivers (as shown in Table 81) would be associated with a higher propensity to save and a lower propensity to consume.

These normal influences were reinforced by wartime governmental policies, such as (1) the encouragement of savings through war-bond sales campaigns, (2) price control which held down living costs, and (3) the rationing of many consumer commodities which aided in keeping spending in check. Added to this was the unavailability of many consumer commodities, such as new automobiles.

The wartime increase in personal income taxes also affected saving-spending activities. But for the steeply progressive tax, it is probable that the savings ratio would have been somewhat higher, and the spending ratio considerably higher. For example, the percentage of total income saved in 1944 was slightly below the percentage saved in 1942, despite the fact that total income was 43 billion dollars higher in 1944. The consumption expenditures of the high-income groups were sharply curtailed by the rise in income taxes, for often these expenditures are reduced while certain savings, such as insurance premiums, are maintained. Table 87 shows estimated saving-spending ratios by income groups for 1942 and 1935-36. It will be noted that the per cent of income spent by the highest income group is sharply reduced in 1942.

Table 87 DISPOSITION OF INCOME, BY INCOME GROUPS, 1935–36 AND 1942

	1935–36		1942	
	Per cent of income		Per cent of income	
	Saved	Spent	Saved	Spent
$10,000 and over	46	42	47	33
5,000–10,000	30	64	36	60
4,000– 5,000	21	73	27	70
3,000– 4,000	16	79	23	74
2,500– 3,000	12	84	20	78
2,000– 2,500	9	87	16	82
1,500– 2,000	5	91	12	86

Source: Consumer Expenditures in the United States, op. cit., and Civilian Spending and Saving, op. cit.

A Restatement

The relationship between the distribution of income and its disposition may be brought out somewhat differently. At some point on the income scale, consumer spending and income are equal (the propensity to consume is unity, i.e., the entire income is spent). For the years 1929, 1935-36, and 1943, this point approximated at a $1,200 annual income. The distribution of income receivers in relation to this point on the income scale is especially important to the amount of aggregate spending and saving and to the spend-saving ratios for the economy as a whole. Table 88 shows, for the three periods mentioned, the distribution of income receivers above and below a $1,200 income and the share of total income distributed on both sides of this line. The years are arranged in terms of the amount of national income, not chronologically.

Table 88 DISTRIBUTION OF SAVERS AND NONSAVERS

	Below $1,200		Above $1,200		
Year	Per cent of income receivers	Per cent of income received	Per cent of income receivers	Per cent of income received	Per cent of national income saved
1935–36	60	28	40	72	10
1929	40	13	60	87	15
1943	25	9	75	91	22

In a year of depressed national income such as 1935-36 the percentage of income receivers in the nonsaving category is much larger than in a

year of high-income levels, such as 1943. Or, in a period of high income the bulk of income receivers will earn enough to permit savings; and the number of nonsavers and dissavers will be relatively small. The effect of this distribution above and below the savings line is seen in the last column of Table 88, giving the percentage of total income saved. For the years covered by this table, $1,200 was the approximate income at which the propensity to consume was unity (the entire income was spent); however, with marked changes in living costs over a period of time, a higher—or lower—income may be the point dividing savers and nonsavers. Thus the greatly increased cost of living in 1947, compared with 1943 and especially with 1935-36, may have raised this figure to $1,800 or $2,000. If so, then the inflationary rise in total money income will not be accompanied by any marked increase in the propensity to save. The evidence indicates this to be true in 1947. In any period some amount of individual income ($1,200, $1,500, $2,000, etc.) will be associated with a propensity to spend of unity; the spending-saving ratios for the economy as a whole will be governed by the distribution of income receivers above and below this point. This distribution normally changes as total income rises and falls.

Emphasis thus far has been placed on (1) changes in total income, (2) shifts in the distribution of income receivers by income groups as total income rises and falls, and (3) the effects of such changes on the saving-spending ratios for the economy as a whole. It is conceivable, however, that the propensities to save and to consume might be altered independently of changes in aggregate income. Three general conditions may be mentioned.

The first is through the redistribution of income. At any given level of national income, a lessened inequality in its personal distribution may be expected to reduce the savings ratio by transferring income from high-saving groups to those who save only a little or nothing at all. On the other hand, an intensification of inequality would tend to increase the savings ratio and decrease the spending ratio. This again emphasizes the bearing of personal income distribution on the saving-spending activities of the economy.

The second is by the application of income taxes. These do not alter personal income distribution but as progressive taxes they alter the distribution of disposable income. Thus an increase in the personal income-tax rates (total income remaining unchanged) would reduce disposable income in the hands of high-income groups and entail a reduction in their saving or spending, or most likely both. The savings ratio for the economy as a whole would tend to be reduced but probably not so much

as is commonly supposed by those who advocate this method of reducing savings, for a transfer of income from the high-income groups goes (1) partly to reduce negative savings at the bottom of the income scale, (2) partly to add to the small savings of those just within the savings category, and (3) partly to increase consumption at the lower end of the income scale. The resultant increase in consumption might exceed the reduction in consumption on the part of the high-income groups, but probably not by a large margin. On the other hand, a decrease in progressive income-tax rates is likely to increase the savings ratio.

The third way in which saving-spending ratios may change for the economy as a whole is through a change in consumer spending-saving behavior. As already seen, spending tendencies were curbed during the war, and savings increased partly by voluntary action on the part of income receivers. In 1946-47, on the other hand, an opposite change occurred when consumers as a whole sought to satisfy demands deferred during the war years.

In general, personal income distribution bears directly on the level of national output, income, and employment. It was shown earlier that the aggregate expenditures of consumers, investors, and government comprise the aggregate demand for the national output. Given total income, the size of the consumer component is determined largely by the personal distribution of income; it is this too (i.e., the propensity to save) that establishes the amount that must be invested if total income in the period ahead is to be maintained. This is to say only that spending and saving activities are among the influences governing the level of national output.

Whether, in relation to its resources, an economy "spends too much and saves too little," or "spends too little and saves too much," or spends and saves the right amounts, cannot be determined a priori. In 1946-47 the propensity to consume, the willingness to invest, and governmental outlays were such that the productive capacity of the nation's resources was taxed, and price inflation ensued. That is, aggregate demand increased while physical production was virtually at maximum levels; as a consequence price increases occurred which raised the money value of national output to the amount of the aggregate demand. Had savings been larger, and consumer demand correspondingly smaller, the general price rise of 1946-47 might have been averted or moderated. In contrast to this period of "overspending," the depression of the 1930's is often characterized as a period of "oversaving," when the expenditure of a larger portion of total income would have increased aggregate demand and therefore national output, income, and employment.

CONSUMER EXPENDITURE AND INCOME DISTRIBUTION

The influence of income distribution on the aggregate amount of consumer expenditure has been examined in the preceding section. There remains the relationship between income distribution and the composition of the aggregate consumer expenditure. By composition is meant the various main types of goods and services for which expenditure is made.

In 1929, 27,000,000 families spent approximately 62 billion dollars for all consumption purposes.

Table 89 COMPOSITION OF CONSUMER EXPENDITURE, 1929

	Billions of dollars	*Per cent of total*
Food	17.0	27
Housing	18.0	29
Clothing	8.4	14
Other living expenses	18.5	30
Total	**61.9**	**100**

Source: America's Capacity to Consume, op. cit.

The "other living expenses" include education, recreation, travel (including automobile), medical costs, and miscellaneous expenses. The composition as well as the amount of the aggregate outlay revealed here are, as always, derived from the spending habits or patterns of consumer expenditure of the nation's families in that year. In some respects these patterns differ greatly by income groups—the poor and the rich not only spend different total amounts but also relatively different proportions of their income for certain types of goods.[7] Table 90 shows for three income groups (1) the total amount spent by each group for large classes of commodities and

Table 90 COMPOSITION OF CONSUMER EXPENDITURE, BY INCOME GROUPS, 1929

Income group	*Food*		*Housing*		*Clothing*		*Other living expenses*	
1. Highest tenth $4,600 and over	$2.9 billions	14%	$6.6 billions	30%	$2.6 billions	12%	$9.5 billions	44%
2. Third tenth $2,450–3,100	2.1	31	1.9	28	1.0	15	1.7	26
3. Eighth tenth $950–1,250	1.3	43	0.8	27	0.4	13	0.5	17

Source: Brookings, op. cit.

[7] See Chap. 13, pp. 288-89.

services, and (2) the per cent of their total income going for these purchases.

Striking differences in individual spending patterns are disclosed in this table, particularly in food and "other living" expenses outlays. The highest tenth spends more in all categories than do the other groups, but only 14 per cent of its expenditure goes for food, compared with 43 per cent for the eighth tenth. On the other hand, expenditures for education, recreation, travel, medical attention, etc., account for 44 per cent of the expenditures of the highest income group, but only 17 per cent of the expenditures of the low-income group. It is evident that income differences impose marked differences in consumer spending patterns.

If income distribution changed, the composition of total consumer expenditures would probably change too. Suppose, for example, that the incomes of the highest group were reduced so that their consumer expenditures declined by 10 million dollars. And suppose further that 10 million dollars were added to the incomes of the $950-1,250 group. What effect would such a change have on the types of consumer expenditure? Food expenditures in the highest group would decline 1.4 million (i.e., 14 per cent of the 10 million dollars) and would increase 4.3 million in the lowest group shown, for a net increase in food expenditures of 2.9 million dollars. Other living expenses among the top group would decline 4.4 million and increase 1.7 million in the lowest group, for a net loss of 2.7 million. On the other hand, if income were shifted from the low to the high group, there would be a net increase in other living expenditures, a net decrease in food expenditures. The other items in this list would not be much affected by a change in income distribution.

Thus it can be concluded that for *any given volume* of consumer expenditure the composition (or types) of expenditure is governed by (1) the patterns of consumer expenditure by income groups, and (2) the distribution of income receivers by income groups. When the distribution of income receivers alters materially, the composition of expenditure will change. This is seen in the 1935-36 estimates.

Total consumer expenditure in 1935-36 was considerably below 1929; and as already noted, income receivers in large numbers had shifted down the income scale during the depression period. However, the patterns of consumer expenditure by income groups in 1935-36 were similar to those of 1929. Consumers in each income group not only spent about as much of their incomes as was spent by the same income groups in 1929, but they spent their money in about the same way. Thus changes in consumer expenditures from one year to another for the main items of goods and services appear to be due primarily to changes in the distribution of income

receivers by income groups. Because of the downward shift of consumers, relatively more of total expenditures went for the purchase of food in 1935-36 than in 1929. The changes in the composition of aggregate consumer expenditure between these two periods are shown in Table 91.

Table 91 CHANGES IN COMPONENTS OF CONSUMER EXPENDITURE

	Per cent of aggregate consumer expenditure	
	1929	*1935–36*
Food	27	34
Housing	29	32
Clothing	14	12
Other living expenses	30	22
Total	**100**	**100**

Source: America's Capacity to Consume, op. cit., and Consumer Expenditures in the United States, op. cit.

Statistical evidence on patterns of consumer expenditure by income groups, and on changes in the distribution of income receivers, leads to the following general conclusion. When total national income increases and income receivers move up the income scale, all components of aggregate consumer expenditure increase, but not in the same degree. Food expenditures increase less than do total expenditures, or total income; that is to say, *income elasticity* of food is relatively low.[8] On the other hand, expenditures for recreation, education, travel, and a host of other things that further the comfort, amusement, and perhaps cultural development increase more rapidly than do total expenditures or income; i.e., these items have a relatively high income elasticity. These changes in consumer expenditure, in turn, bring about corresponding changes in the composition of national output when total output rises. When total income falls, and with it consumer expenditure, expenditures for food decline more slowly than total outlays, and expenditures on education, recreation, etc., decline more rapidly.

Tax policy may affect patterns of consumer expenditure in important ways. Steeply progressive income taxes have the effect of reducing inequalities in *disposable* income and therefore in consumption. Many large country estates and mansions in the cities have been abandoned by high-income recipients who are no longer able to afford their maintenance after settling with the tax collector. In other respects, too, consuming habits of the wealthy

[8] See Chap. 13 for a discussion of income elasticity.

have been altered by the reduction of their disposable incomes by taxation.

This discussion of the components of consumer expenditure has been confined to the very large types of expenditure. A more detailed analysis of the subtypes within these larger classifications would reveal still greater differences in the spending patterns of consumers, especially as they are magnified from prosperity to depression. But such detail is endless. The discussion given here has established the following general points:

1. Patterns of consumer expenditure differ widely by income groups.
2. The composition of aggregate consumer expenditure is largely a function of the distribution of income receivers by income groups.
3. From this it follows that as income receivers shift up or down the income scale with the rise or fall of total income, the composition of consumer expenditure undergoes a change.
4. As long as consumer spending patterns by income groups remain relatively unchanged, a redistribution of a given volume of national income would also change the composition of aggregate consumer expenditure.

CONSUMER EXPENDITURES AND THE ALLOCATION OF RESOURCES

The foregoing discussion can now be brought to bear on the allocation of economic resources—labor, capital facilities, and natural resources. Consumer expenditure is especially important here because not only does it guide the composition of consumer goods and services (the largest part of the national output) but it also directly influences the decisions on what types of capital goods will be produced. In short, the *pattern of production,* and therefore the *pattern of resource use* (or allocation of resources), are largely determined by the amounts and types of things consumers will buy.

The amount of each type of consumer good currently produced depends upon the short-term profit expectations of enterprise. Producers attempt to judge the demand for their goods in making their production plans. The demand of the individual and the market demand for each type of consumer goods are measured by, or expressed by, the quantities purchased at given prices. The quantity-price relationship determines whether production is profitable or not.

Therefore, consumer demand or expenditures—as modified by producer selling effort—*selects* the goods and services, and thereby the industries, that are to survive. The pattern of consumer goods production is in this way determined largely by the composition of consumer expenditure. As

the composition changes in a short period of time (because of changes in national income and its distribution) the pattern of production shifts too.[9] Over long periods of time—with the growth in the level of national income, a rise in the average family income, the introduction of new products, and the changing costs of many old products—the composition of consumer expenditure (consumer demand) undergoes change and with it the pattern of production.

Not only does consumer expenditure influence the pattern of consumer goods production, but it exerts in normal times a dominant influence on the pattern of capital goods production as well. Capital facilities are purchased by the producers of consumer goods on the basis of their long-term expectations of profit; again, this involves a judgment on the cost-quantity relationship in the future, or an estimate of future consumer expenditure for given types of goods or services. Thus the consumer goods industries invest in capital facilities on the basis of probable long-term consumer demand or expenditure; [10] in this way expected consumer expenditure influences both the amount of capital goods output and the pattern of its production.

Since the production patterns are largely guided by the composition of consumer expenditure, income distribution has an important effect on resource allocation, for income distribution can change the composition of consumer outlays in many ways. It may be objected that production and productive capacity are not closely meshed with consumer demand; some industries may suffer from considerable overcapacity which stands idle a good part of the time. Such disjointedness is readily evident. Economic resources are not necessarily employed in the only possible or the most orderly fashion, but the fact remains that the economy depends upon its market decisions ultimately—it produces what is demanded effectively. What people want and spend their incomes for depends upon many things, but chief among these is the incomes at their disposal. These incomes, as distributed, govern in general what income groups are able and willing to buy and, therefore, the way in which economic resources are used or allocated in the economic system. If in the long run the economic progress of the nation leads to an upward shift of incomes on the income scale, or if social influences effect a change in the proportionate distribution of income, the patterns of production and the allocation of resources can be expected to change accordingly.

[9] Some shifts, of course, are slow. Food production changes but little from prosperity to depression; on the other hand, great changes occur in the production of manufactured goods as consumer expenditures rise and fall.

[10] Taking into account probable long-term substitutes for capital and complements of capital facilities.

CHAPTER 39

The Determinants of Income Distribution

The causes and conditions underlying income distribution have long been a subject of interest and controversy. The existence of income inequality in particular has raised many searching questions about the distribution process and the moral, social, and economic consequences of its operations. Why is there a marked inequality in the distribution of incomes? What, if anything, should be done about it? These questions are as old as civilization. The division of society into the rich and the poor in the past called forth laments from the Prophets and denunciations from the Christian Fathers. In recent times opposition to this division has been the inspiration of political platforms and appeals to violence from reformers and revolutionaries. Greed and avarice explained the existence of great wealth to the satisfaction of the Schoolmen: by their reasoning, honesty and devotion to spiritual affairs accounted for poverty. Moralizing about income distribution, however, is by no means confined to the Age of Faith. The modern socialist doctrine of "exploiters and exploited" is a secular version of the ancient injunction against the triumph of greed over honesty and industry. Moralists have not agreed, however, in their verdict on income inequality; those of another persuasion have viewed wealth as the reward of the Christian virtues of industry and frugality, and poverty as the wages of the sins of indolence and vice.

ASPECTS OF INCOME DISTRIBUTION

Many individual and social interests are involved in the income-distribution problem. The spiritual well-being of man is the concern of some people, the material well-being of man is the concern of others; and in recent times many have placed an economic emphasis on distribution as

it affects the functioning of the economy through the saving-spending relationship and the disposition of income. Whatever the interest in the matter may be, the question recurs: How is income distributed, and what principles underlie the sharing of income? There are a number of possible ways a community or economy can share its income.

Income Equality

It is conceivable that a society could provide actual equality in distribution, as some reformers have proposed. This might mean an equal income for each family in the economy, or since family size differs, for each person. Although this equalitarianism is given little serious thought as a national policy—except possibly by some whose current income compares unfavorably with the average—it is a standard frequently invoked in efforts to reduce inequality. Equality has a strong appeal when limited to a narrow range of income problems. "Equal pay for equal work" is a slogan of importance in feminist as well as in labor-union circles, and union-employer contracts frequently include provisions assuring equality in this limited sense. Similarly the view that "farmers should get as much as factory workers" undoubtedly draws support for farm benefit legislation.

Whatever may be said for equality on other grounds, this principle leaves unanswered some fundamental economic questions. The payment of income is, in a free-enterprise economy, the means of allocating labor and other economic resources among the great variety of uses for these resources. But income *equality* would necessitate some other method of allocation, such as government edicts on all economic processes. In a free system equality would discourage the bulk of private investment—at least as it is carried on under present circumstances. It would tend to reduce the volume of current savings by individuals, and the functions of saving and investment presumably would cease to depend largely on private decisions. In short, income equality would call for a basically different kind of economic organization and—since material incentives and means of allocation would be radically changed—a different kind of social and political organization also.

Income and Need

Nineteenth-century socialist literature gave prominence to the slogan "From each according to his abilities, to each according to his needs." Each person is to work to his best ability, but will be paid in accordance with his

needs. Effort and income are thus dissociated; the income incentive is replaced by a social incentive, the discharge of one's "responsibility to society." This principle if effected would tend to wipe out marked inequality. Income differences would measure differences in "need," as determined by some kind of standard.

Economic questions are raised here also. Income payments based on need would afford no basis for resource allocation, nor the incentives to work, save, and invest. Other incentives and other methods of performing the necessary economic functions now carried on by income payments would have to be established by the social and political organization within the economy.

Actually, of course, need as a basis of income payments is found in the distribution process in this country, albeit on a relatively small scale. Most of the income payments to nonparticipants are related to need.[1] Persons on public relief customarily receive an amount determined by a "needs test." Benefits under the old-age insurance system are larger for a retired man and his wife than for a single man; some other social security benefits as well as veteran payments are related also to the number of dependents.

Similarly, participants may be given at least parts of their incomes on the bases of need. Servicemen's pay varies with the number of dependents. In Canada and in other countries families receive an allowance from the government for an increased number of children. To a small extent some firms pay employees more money if they are married, but this is not characteristic of labor services generally. Perhaps more than any other measure instituted in this economy, the income-tax laws reflect the socialist concept. Need is recognized in the number of dependents in the calculation of personal income-tax liability, and the progressive character of the rates is a special application of the principle "from each according to his ability." Need, however, plays virtually no part in the determination of income from private sources, and to the extent that it figures in nonparticipating incomes it has an insignificant effect on the direction of the economy.

Income and Social Status

Income distribution may be governed by the social class or status in which people find themselves. In the feudal period, income was distributed in this manner: in principle each person had a right to the income which custom allotted to his particular occupation and station in life. The com-

[1] Indeed, within the family income is normally shared among family members on the basis of their needs.

munity was conceived as an organic whole, an ordered hierarchy, with serfs, artisans, merchants, the landed nobility, and the clergy performing essential functions. Each person had the duty to carry out his work and fulfill his station in life and the right to the perquisites—including a customary income—appropriate to his station. Vestiges of this system prevail today in England and on the continent in social if not in political and economic forms. With the rise of fascist governments in Germany and Italy the concept of the organic or functional state was revived as part of their social ideologies, although it cannot be said it played a dominant part in the distribution of income, for both countries retained a form of capitalist economy.

This principle of income distribution is a factor of importance only in a society where all activities are subject to control, by religious authority, a monarchial order, or a political bureaucracy. In such systems income distribution and economic activity in general are subject to customary influences or bureaucratic decree—in either case status tends to determine what a person gets and what he does.

While feudal, or at least manorial, ideas have had relatively little influence in this country, the idea of a "right" to income and a customary income is not entirely absent. Frequently it is said that "a working man should get" a certain income, or that farmers have a "right" to a certain income, or businessmen to a "fair" profit. It is a common practice to maintain executive salaries at levels commensurate with an assumed social position; and it is an axiom that those who work with their hands work for less. The concept of a class society is considered alien to American historical and cultural development—the word "masses" is deplored when applied to American workers—but it is nonetheless true that even in the United States, as persistent economic circumstances lead large groups to accept their status as relatively fixed, income will increasingly be regarded as a "right."

Income and Productivity

Of greatest significance in the capitalist system is the principle that relates income to productivity, i.e., each person receives an income roughly equal to the value or importance of his contribution to the output of the economy. If one service is regarded more highly than another—the service of a physician as compared with the service of a domestic servant—the payment (income) for one will be higher than the other. This importance, in turn, is determined by the relative scarcity of the services in terms of the demand for them.

Personal income distribution effected under the principle of productivity performs a dual function. First, it tends to allocate productive services or resources; high incomes attract resources, low incomes repel them, and thus productive services tend to shift about in accordance with the demand for them. Second, income is an incentive to perform productive services, or to produce needed goods and services; and the incentive varies in accordance with the demand for goods and services. Under the productivity principle, then, income distribution is both an inducement to produce and a means of guiding the use of economic resources.

Productivity may be regarded, too, as establishing the "right" to income. Historically this right was asserted by spokesmen of early capitalism against the income claims and other privileges of the nobility, the church, and the king, whose claims and privileges were supported by custom and social philosophy with little or no regard to productive effort. Why should those who work and risk their capital support those who do neither? In the course of time the "right to the fruits of one's labor" became established.

Acceptance of the principle, however, has not been accompanied by agreement on its operation. How is productivity measured? And who are the producers? Agreement is general that human effort or labor is productive when directed to desired ends, but are the owners of capital and land producers? The prevailing view in this and other capitalist countries is that land and capital are productive and thereby entitle their owners to income. Against this Proudhon, a noted socialist of the last century, asserted that "Property is theft." For Karl Marx only labor was productive, and nonlabor income (rent, profit, and interest) measured the extent of exploitation of labor by the propertied class. The "right of labor to the whole product of labor" thus became a socialist slogan embodying a special interpretation of productivity as the basis of income distribution.

The Bases of Personal Distribution of Income

There are, then, at least four principles which may guide the distribution of income: equality, need, social status, and productivity. Although productivity is the ruling principle underlying income distribution in this country, other influences are at work which modify the operation of this principle. In general and apart from monopolistic practices, these other influences tend toward a reduction of the income inequality which results from differences in productivity as governed by supply-and-demand relationships. Within the pattern of income distribution there are, as has been noted, participating incomes and nonparticipating incomes, the former

based on services rendered to production and the latter on legally established claims made (ultimately) against the former. Political and social factors intervene in this distribution and either potentially or actually modify it. These factors bring about (1) a change in the conditions of supply and demand, either to increase or decrease the incomes of certain groups, and (2) a series of programs that pay incomes to certain groups and take incomes from other groups. The first type of governmental intervention includes agricultural price and production policies, minimum-wage legislation, and encouragement or discouragement of union organization. The second type largely embraces progressive income taxes and the relief and social security programs.

INCOME INEQUALITY AND OWNERSHIP OF WEALTH

Income differences, as has been indicated, stem largely from the unequal distribution in the ownership of income-bearing assets, and from differences in the economic importance of personal services. The assets which yield income to their owners may be tangible property, such as buildings, machinery, equipment, land, and other resources; or they may be corporation stocks which are evidence of ownership; or they may be evidence of indebtedness, such as bonds and mortgages. Approximately 20 or 25 per cent of the total national income is received by virtue of ownership of such income-bearing assets. Ownership is heavily concentrated in the hands of a relatively small segment of the population, and accordingly the income from these assets is highly concentrated. An estimate for 1927 showed the distribution of wealth in the United States as presented in Table 92.

Table 92 ESTIMATED DISTRIBUTION OF WEALTH, 1927

Class	*Per cent of population*	*Per cent of national wealth owned*
Richest	2	40
Middle	33	45
Poorest	65	15

Source: W. I. King, "Wealth Distribution in the Continental United States," *Journal of the American Statistical Association*, June, 1927.

The concentration of ownership indicated by Table 92 does not reflect the concentration of property income (e.g., rent, interest, dividends) because the assets of the rich and the upper middle group are usually the highest income yielding assets while the wealth of the poorest group may

be nonincome or low-income bearing assets. Other estimates also show the marked concentration of wealth. In 1891 the poorest 65 per cent owned less than 5 per cent of the wealth of Massachusetts and Wisconsin, and the richest 2 per cent owned nearly 60 per cent. Virtually the same degree of concentration of wealth existed in France and Prussia in 1908-09 as in these two American states. The United Kingdom in 1908 showed an even greater concentration; the richest 2 per cent owned over 70 per cent, the poorest 65 per cent about 2 per cent of the wealth.[2]

Some idea of the relationship between high income and the ownership of wealth can be gleaned from the figures of the Bureau of Internal Revenue in Table 93, showing the proportions of income derived from personal services and from property in 1930 for several income classes.

Table 93 SOURCES OF INCOME, 1930

Income class	Per cent of income from personal service	Per cent of income from property
Over $1,000,000	2	98
$25,000–50,000	30	70
5,000–10,000	62	38
1,000– 2,000	82	18

Source: U.S. Treasury Department.

It is evident from these figures that income inequality would be much less pronounced if (1) all incomes were derived from personal services alone, or (2) if income assets were evenly distributed—or at least less unevenly distributed. The major difference between the incomes over $1,000,000 and the incomes of $1,000-2,000 is found in the proportion of income derived from property. The importance of wealth or property in accentuating income differences has led to a number of proposals to nationalize wealth or break up large individual holdings. "Land nationalization" and "redistribution of land" have been popular slogans in countries where a relatively small group of landlords owned a large part of the land—England, Poland, prerevolutionary Russia, Mexico, China, and other countries. Such proposals have never had much popular appeal in this country because of its large amounts of land, its increasing landowning population, and its free market in land. Moreover, land is not so important a source of income here as in other countries; great wealth in this country is chiefly industrial and commercial.

The concentration of wealth is fostered by inheritance. Some of the

[2] Estimates of W. I. King, quoted in Fairchild, Furniss, and Buck, *Elementary Economics*, Macmillan, 4th ed., 1939, Vol. 1, Chap. 21.

present-day fortunes in America trace back to the early nineteenth century. In many older countries wealth in the form of land has passed from generation to generation for centuries. In the historical process fortunes may be increased, for the large incomes from wealth provide the means of acquiring more—wealth begets wealth. To prevent this pyramiding and consequent accentuation of income inequality, governments in many countries have introduced steep death taxes. Wealth may be passed on but in transit the state lays claim to a substantial part of it. At the same time high incomes are heavily taxed, thus limiting the possibility of adding greatly to an existing fortune or initially accumulating a great fortune. Such tax measures are a product of sociopolitical influences aimed at a partial redistribution and lessening of income inequality.

INCOME INEQUALITY AND PERSONAL SERVICES

If private ownership of *income assets* were abolished, or distributed evenly throughout the population, the spread between high and low incomes and the concentration of income would obviously be lessened. Income differences nevertheless would still be substantial, and the proportionate distribution of income would still reveal a considerable concentration because of differences in the prices of personal services. Personal services command different rates: a clerk in a country store may get $600 a year, the head of a large corporation, $600,000. The services of both are "productive services" but in the market for productive services the importance or productivity of one is rated one thousand times greater than the other. How can such differences be accounted for? Can the services of one be one thousand times more important than those of another? Both men would probably give different answers to the second question, but the fact remains that the market assigns this difference; these are the prices offered for the particular services. Worth or importance or productivity may be measurable in many ways, but the market has only one method: the inherently variable payment of prices or incomes as governed by demand-and-supply relationships.

The Demand for Personal Services

Several aspects of the demand for personal services may be mentioned which help explain great differences in income. The public fancy for certain forms of radio entertainment and for certain movie personalities may be taken as special examples of the peculiar quality and intensity of de-

mand which results in pronounced inequality. The demand for songs, drama, and comedy tends to be concentrated on certain types of songs and singers, particular screen personalities and comedians. The demand may be fickle, but current favorites enjoy an enormous popularity (and large incomes) while it lasts. Each enjoys a monopoly position, each provides a differentiated service—a voice or figure or face or a type of radio personality—adequately and skillfully developed by the most effective publicity techniques. It is true that the supply of such talents is limited by natural endowment, among other things, but the character of the demand facilitates such limitation of supply. The "public" wants a favorite, or a small number of favorites, not a large group on which to bestow its attention and affection; and it is often true that the favorite is no more gifted than many nonfavorites.

The personal service by such performers directly satisfies the wants of the public constituting the demand. Strictly speaking, radio listeners do not demand these services in the sense of paying for them; the demand comes from the manufacturers of soap, toothpaste, or cigarettes who sponsor the programs and who presumably know what the public wants for entertainment.

The demand for the services of wage and salary workers by business enterprisers, and the demand for the services of farmers and other entrepreneurs is a *derived demand,* derived from the ultimate demand of consumers for the things produced by these services. For wage and salary workers "demand" means the ability and willingness of the employer to pay a given wage or salary. Demand in this sense may affect incomes differently, or be subject to influences which accentuate income differences.

For example, the executive of a large corporation may receive $600,000; the demand of the corporation, i.e., the board of directors, for the special capacities and experience possessed by the executive is measured by this large sum. But the demand in such a case may be inflated, in that the executive would have taken as little as $200,000, or someone else equally as capable could have been hired for this lower amount. Why, then, does he receive the larger amount? In the realm of higher corporate bureaucracy the demand for executive services is often determined by the men who provide the supply of these services. The board of directors in this case may be composed of men holding prominent executive positions in banks and other corporations who themselves receive large salaries and bonuses. The $600,000 executive himself may be a director of several other corporations. To cut the "going rate" of executive salaries in one important corporation may impel cuts elsewhere. Since those who determine the "de-

mand" for executive services generally also compose the "supply," there is a community of interest which may be expressed as rigging the market. A market so organized enables those who sell their services there to command income advantages not found in other markets. The wage earner employed by the same corporation would normally have no similar influence on the demand for his services. To be sure, if he were a union member he may obtain a wage higher as a result of the union's partial control of labor supply. But no community of interest exists in this case; the worker may get somewhat more than he would be willing to take (if he had to) because of the coercive power of the union. The executive gets much more than he would be willing to take if need be, i.e., if he were treated by the board of directors in the impersonal way that he himself treats the ordinary wage earner far down the line.

But the contrast here is not entirely explained by a rigged market in one instance and a free market in the other. If both markets were equally impersonal the executive would still get a greater income than the wage earner by virtue of his special capacity or importance or productivity, i.e., the scarcity of his ability. The point to be made is that income inequality is exaggerated by such market differences. Indeed, the character of the demand for personal services is more diverse than is indicated by the examples here. At one end of the scale, the demand for and supply of services are virtually identified in one and the same person, as when the principal owner of a company installs himself as executive and fixes his own salary as well as perhaps the salaries of relatives and friends. The demand for executive services already described previously is, of course, a less extreme control. At the other end of the scale, there is a large supply of persons who have no personal influence on the demand for their services. The employer assumes a "take it or leave it" attitude, and the individual in such a market is without recourse. While the position of the employer is stronger by far than that of the individual worker, his power is not absolute, for he is in competition for labor with other employers and will generally pay the rate of alternate opportunities for workers. Even this qualification, however, is mitigated when there is an oversupply of labor in a market or if there is temporary agreement among employers to keep rates low.

The Supply of Personal Services

The demand for each type of service is governed in general by the productivity or importance of the service to those who want it. Productivity in turn depends upon the supply of the service in relation to the demand for

it. When, in relation to demand, the supply of a service (or type of labor) is large, the value of its marginal product is relatively low; when a service is scarce, the value of its marginal product is relatively high. The principle of diminishing marginal productivity has been described in an earlier chapter.[3] Thus from an income standpoint, the difference between a skilled patternmaker and a common laborer is one of relative scarcities.

There arises now the question: What factors govern the relative scarcities of different services? If the great scarcity of brain surgeons and competent executives enables persons in these lines of activity to command large incomes, why are there not more brain surgeons and able executives? If common labor is poorly paid because its supply is relatively large, and the value of its marginal product therefore low, why are its ranks so amply supplied with new recruits year after year?

Personal Qualities

Differences in natural endowment, for want of a better term, are a partial explanation. Individuals differ in learning capacity, interest and ambition, artistic inclinations, in qualities of leadership, in personality and character. Professional work, for example, requires a learning capacity of a higher order than does unskilled manual work. Ambition, interest, and a willingness to submit to discipline and training are required in greater measure of the skilled than the unskilled trades. Such personal qualities and capacities are unevenly distributed; this fact alone is sufficient to limit the supply of those personal services which require them.

Ability for present purposes must be regarded as *money-making* ability. Thus some lawyers and physicians stand well above their colleagues financially, not always because of superior professional competence but because of a "good business sense" and a keen eye for the profitable aspects of legal and medical work. Money-making ability need not be associated with intellectual, artistic, or even manual abilities. A man may by normal standards be gross, stupid, and awkward, yet possess the shrewd sense that enables him to find a place high on the income scale. Personal capacities and interests thus vary enormously. The market for personal services assigns high prices to certain types in limited supply, lower prices to those more commonly found in the bulk of the population.

[3] See Chap. 17.

Opportunities

The opportunities provided by the environment in which people live are important also in governing the personal distribution of income. Opportunities may enable a person to become qualified for remunerative work—such as special training and the cultural background that may be requisite—or they may relate to the accessibility of suitable work and of business prospects.

Opportunities of these types are not equally available to all persons and apparently have little direct relationship to the endowment of native ability. The unequal distribution of opportunity is caused in good measure by unequal distribution of wealth and income. The children of the poor have fewer opportunities than do those of the rich to obtain the education, training, and general background necessary for many remunerative employments. The son of a wealthy family often has a position awaiting him, and his advancement is often made easier by his connections. Moreover, he is more likely to know where favorable income prospects lie, and to be financially able to move in search of them. The paucity of opportunity confronting one who lives in a city slum or in an isolated rural community is often offset by a person's ability and ambition to get ahead, but the native ability and the ambition usually must be of a higher order if this person is to compete on equal terms with one whose environment is more conducive to financial success.

A word should be added on the importance of luck. Even with reasonably similar social environments and native abilities, income differences may be substantial. Chance may lead one person into an occupation which later becomes unusually remunerative, or which opens the way to another unexpectedly profitable opportunity. Luck was with those who, many years ago, chanced a few hundred or a few thousand dollars in the securities of a small company producing a virtually unknown product—Coca-Cola. Many such instances may be found—and many more when luck looked the other way.

Noncompeting Groups

Personal qualities and the opportunities afforded by the environment exert a selective influence on the supply of personal productive services. Moreover, these factors tend also to limit the shift of people from one occupational category to another. A shortage of doctors cannot be met by

volunteers from the legal profession or from the ranks of industrial workers. Even if the lawyer had all the necessary native endowment and interest required for the medical profession, he lacks the years of training. Moreover, each professional group and skilled trade places obstacles in the way of such shifts. The professional associations jealously guard against infiltration by outsiders. Standards of competence are of course maintained, but at the same time a relatively high degree of scarcity is assured. Many trade unions practice a policy of exclusion to keep services relatively scarce and thereby protect their membership. These restrictive measures limit opportunity in varying degrees, and make certain employment noncompeting, i.e., relatively free from outside competition. Only the great groups of more or less unspecialized productive services—the unskilled and semiskilled industrial worker, the office-worker groups, and similar segments of the labor force—lack the protection afforded by an accrediting association, or apprentice regulations and high union dues, or long years of study and training; as a consequence they are concentrated on the lower end of the income scale.

CONTROL OF SUPPLY AND DEMAND

The fact that scarcity is remunerative has led to a widespread movement on the part of occupational groups to gain control over the supply of their respective services. This movement goes far back into history; the ancient restrictions against the spread of knowledge may have been partly motivated by pecuniary considerations. The merchant and craft guilds of the Middle Ages imposed restrictions on entrance into the guilds and upon outsiders in part, at least, for financial protection. The modern era has witnessed a recrudescence of the guild idea in many ways and largely for the same purpose: to assure the memberships, whether they comprise professional groups, wage earners, or businessmen, of a favorable income position. Success has crowned the organizational efforts of those groups whose membership would be, in any case, relatively small; but many groups, such as the unskilled and semiskilled, have had difficulty in organizing the supply of their services.

The average income of any group may be increased in one of two ways: (1) by restricting the supply of personal service in relation to the demand for it and (2) by increasing the demand for personal service in relation to its supply.

Regulation of entrance into an occupation is exemplified by the training qualifications needed for a person to practice law or medicine; these are largely determined by the professional associations and incorporated into

the laws of the states. Similarly, some labor unions seek to control entrance into the trade by requiring prospective members to serve an apprenticeship. An important advantage of union control results when employment is open only to members of the union. Entrance into an occupation may also be regulated by the public requirement of licenses, which result in restricting the number of liquor dealers, cab drivers, public-utility enterprises, and certain other forms of enterprise. Although such licenses are issued partly to protect the public interest from undesirable (e.g., criminal and incompetent) enterprise they serve also to restrict entrance for the (income) protection of those already engaged in the occupation. Regulation of entrance into an occupation may therefore be achieved by the power of a union, association, or other form of organization, or by government, in order wholly or partly to secure a favorable income by limiting supply in relation to demand.

Some types of labor legislation directly influence the size of the labor supply as a whole or the size of the supply available for certain kinds of employment. Federal and state child-labor legislation controls entrance of young persons into the labor supply. State laws limit the employment of women under certain conditions, thus restricting the supply of labor for industries which largely employ women, such as the garment industry. Indirectly, old-age relief and old-age insurance restrict supply by encouraging the retirement of workers after sixty-five. Such measures restrict a quantitatively small segment of the potential labor supply, but they remove some competition and thereby tend to raise the rates of wages in certain occupations.

Control of supply can also be effected by regulating or at least influencing conditions under which personal services are supplied, without necessarily regulating entrance into an occupation. Thus, while a union may not exclude anyone from membership it may try to impose on the employer certain wage rates and working conditions at which union members, as well as possibly nonunion workers, shall work. In effect unions seek to establish supply prices for the services of their trade. Governmental establishment of minimum wages and other wage requirements is another important instance of controlling the conditions under which personal services are supplied by setting minimum supply prices.

Many governmental measures are intended to influence incomes by directly or indirectly affecting the conditions of supply of personal services; others seek to affect incomes by influencing the supply and therefore the prices of commodities.

The first type of legislation is represented mainly in labor laws. The National Labor Relations Act of 1935 affirmed the right of labor to

organize, required the employer to bargain with unions, and forbade employer interference with union organizational efforts. This legislation encouraged the growth of unions; in 1930 only 10 or 12 per cent of the nation's wage earners were organized and in 1947 about 40 per cent. Labor unions, thus encouraged and protected, have grown into powerful factors determining the conditions of supply of labor (and therefore wage rates) in large sectors of the American economy.

Government determines the supply prices of services to an extent when it enacts minimum-wage legislation. The Fair Labor Standards Act of 1938 raised the hourly wage rate of an estimated 10 million workers; its provision for overtime premiums benefited many additional workers. The Federal government requires the payment of prevailing wages (in effect, union wages) by employers working under government contracts. Also, most states have legislation affecting wage rates, either for workers employed on state projects or for industries not subject to federal law. The purpose of such legislation is to fix a minimum supply price for labor above the level which prevails under the free operation of supply and demand in the markets affected.

The second type of legislation is found extensively in agriculture, where government intervention is focused on the supply of and demand for the commodities produced by the services or efforts of farmers, rather than on the services themselves. In the prewar decade the volume of agricultural commodities produced was large in relation to demand, resulting in a low aggregate and average farm income. Agricultural policies of the New Deal administration attempted to "adjust" supply in relation to demand; in effect, this meant a planned reduction of agricultural production. Because the demand for farm products is relatively inelastic, a reduction in output was expected to bring about greater than proportionate increases in farm prices and therefore an increase in farm income. The purpose was to attain "parity" prices and parity income.[4] This latter term emphasizes the income objective of the policy; parity income was defined as the relationship existing between farm income and nonfarm income during the base period 1909-14. During the 1930's farm income was below parity, i.e., in terms of the previous relationship farmers were getting less than the nonfarm group.

The limitation of the supply of commodities is also a purpose underlying tariff policy. By restricting the importation of foreign goods through duties,

[4] See Chap. 28. The effectiveness of this policy is difficult to assess. Price and income increases occurred in the years following the adoption of this policy. However, farm production was sharply reduced by droughts in 1934 and 1936 and the rise in price due to this might have outweighed the rise owing to policy.

the domestic producer is enabled to sell his output at a price higher than would otherwise prevail or to capture a larger part of the market. In either case the protected producer is expected to increase his income or to be saved from an impending loss of income. As noted in Chapter 36, the United States has long been a high-tariff country; this form of intervention in income distribution, therefore, has been widespread.

The examples given here are sufficient to indicate the extent and types of governmental intervention in regulating the supply price of services directly, strengthening organized groups to enable them to influence supply price, and regulating the production and marketing of goods to strengthen the income position of the producers.

Influences on Demand

Both government and organized groups may regulate entrance into an occupation or influence the conditions of supply by other means, but the success of such measures and actions rests ultimately upon the demand for the services thus regulated. For example, unions may restrict entrance or otherwise increase the wages of groups of workers, but if the demand for the services in question is not increased the wage rise will be accompanied by a fall in employment. Or a minimum-wage law may have the effect of causing low-productivity workers to lose their employment unless the demand for their services increases. Government can forbid the employment of workers below a minimum wage, but it cannot require the private employment of labor at or above this wage. If the wage exceeds the marginal product of labor, employment will not be offered. Thus the regulation of conditions of supply does not assure the purpose, namely, the increase in the income of the groups in whose behalf the control is instituted. In some cases, governmental measures regulating the supply price of wage earners, or the incomes of other groups such as farmers, are accompanied by actions that assure the payment of the income. Thus when government establishes union wage rates as the minimum to be paid by employers working under government contracts, the contract price is sufficient to assure the payment of these rates; in effect, the government determines the supply price of labor and the demand for labor at that price. Similarly, when the government establishes parity price as the minimum at which crops may be sold, it makes this policy effective by buying any part of the crop not purchased by the trade at that price. Trade-union rules sometimes provide for the maintenance of employer demand for wage earners; the "stand-by" and the helper often constitute needless employment, and some of the "feather-bed"

rules of certain railroad unions accomplish the same purpose. In general, however, an increase in the supply price of labor through unionization and governmental wage policies does not *directly* increase the demand for labor by employers.

The question now arises: Is there an *indirect* connection between a wage increase, as discussed here, and a subsequent increase in the demand for the labor subject to the rise in wages? Since the principal demand for labor, or personal services, arises from the demand for consumer goods and producer goods, the question can be restated as: Does such a wage increase augment the demand for consumer goods? for producer goods?

Under some circumstances, wage increases might augment consumer demand, but this effect is not certain. Consumer demand would increase if (1) the wage rise shifted income from the high-income brackets (i.e., the wage increase is not offset by increased prices or a drop in employment) and (2) idle resources were at hand to permit an expansion of real output. A measure of income redistribution as assumed here would tend to increase the spending ratio for the economy as a whole; a rise in consumption would, by means of the acceleration effect, tend to increase investment with its multiplier effect. Thus, given some income redistribution, and the presence of idle productive capacity, there would ensue an increase in national output and the demand for personal services *as a whole*, including those subject to the initial wage increase.

The circumstances assumed here, however, are not likely to obtain in practice. When economic resources are at relatively full-employment levels, an increase in wage rates is likely to be passed on in higher prices. Although the demand for the wage earners subject to the wage rise might increase (i.e., employment might be maintained at the increased wage), the wage increase would not mitigate income inequality. If the demand for the products produced is relatively elastic, the increased wage cost might not be passed on in higher prices; in this case the rise in wages would probably be offset in time by an increased use of capital and decreased employment in the occupations affected. Both union actions and governmental measures have brought about substantial increases in wage rates, but the evidence suggests that little or no redistribution has been effected in this way, i.e., the ratio of wage income to total national income appears to be unaffected by such actions and policies.[5]

[5] See Chap. 29.

Other Supply and Demand Influences

For the groups discussed thus far—principally wage earners and farmers—the sociopolitical influences at work on the supply-demand influences seek to raise low incomes and thereby lessen income inequality. This general trend is manifested in other forms of intervention; e.g., in the antimonopoly policy of government. Hesitant as it is, its purpose is to break up restrictive practices and establish competitive practices. In a sense, this may be regarded as an increase in the supply of personal services in the industries in question, for breaking up monopoly usually (but not always) paves the way for new entrants into the field. This, in turn, tends to reduce high incomes in the industries in question. Closely related to this is the attempt on the part of government to encourage small business by providing business information, financial and other aids, in order to stem the growth of big business.

While these and other influences work generally toward the lessening of income inequality by intervening in supply-demand relationships, the pattern is by no means a consistent one throughout. For example, the tax policies of government tend to limit the expansion of small business without reducing the scope of big business. Governmental support of unionism tends to curb the growth of small firms but not of large firms. Restriction of agricultural output under governmental auspices is accompanied, apparently incongruously, by governmental research work for the purposes of increasing agricultural output. Antimonopoly is a policy of government, but its execution has generally been lethargic; and, indeed, tariff policy tends to promote monopoly.

Thus governmental intervention in supply-and-demand relationships for the purpose of reducing income inequality is not clear-cut. Policy is piecemeal, subject to the political pressures and economic exigencies of the groups who look to government as an ally in the struggle to get more income. Nevertheless, the general direction of the major policies affecting income distribution by modifying supply-and-demand relationships is toward raising the low-income groups and minimizing income inequality. When the policies described here are viewed in conjunction with others that affect income distribution, such as corporate and personal income taxes, inheritance taxes, unemployment insurance and public works, old-age insurance and public assistance, the conclusion offered above is strengthened considerably. The following chapter deals with some of these major policies, particularly those that provide income to groups unable to earn a minimum income through their own productive efforts.

CHAPTER 40

Governmental Assurance of Income

Direct intervention by government in the personal distribution of income is represented in this country mainly by the large social insurance and relief programs established. Large-scale assurance of income to persons whose normal income is cut off began during the depression of the 1930's, when extensive programs of unemployment relief, public works and work relief, unemployment insurance, and old-age insurance, were established. Approximately 10 per cent of the nation's income receivers annually obtained income from these programs during the prewar decade; the expenditures averaged about 3 billion dollars a year. From the emergency measures established in this period grew the principal long-range social insurance programs presently in effect. Most of the wage and salary workers are presently eligible to receive unemployment insurance in case of unemployment and compensation benefits in case of injury while at work. The bulk of the gainfully employed—nearly 50 million of them—are engaged in employments covered by the old-age insurance system, which provides monthly annuities to contributors after they attain age sixty-five, and survivor benefits to certain dependents of contributors who have died. Various forms of local, state, and federal relief are also available to persons in need, most of whom are not benefited by the other income assurance programs. In the prewar decade, furthermore, large government work programs provided public jobs for unemployed workers.

Although local and state relief on a limited scale has been a feature of this economy from an early period, and workmen's compensation for industrial accidents made its appearance in some states before World War I, these programs were unable to alleviate the hardship of unemployment following 1929. The inadequacy of these means and the social reform predilections of the 1930's caused to be established not only emergency relief and public employment programs but also the permanent means of assuring

"income security" through social insurance. A feature of this development is the large role of the Federal government in establishing these programs and its financial and administrative participation in them. Indeed, this development may now be called a "movement" toward governmental income assurance, so deeply is it imbedded in federal and state legislation and in the public consciousness.

Similar developments in other countries, particularly in Great Britain and in some of the industrial nations of Europe, have carried programs of governmental assurance of personal income considerably farther than in the United States. These other countries embarked upon this course earlier than the United States and the momentum of this movement has carried them further toward assuring the bulk of their populations against a complete loss of personal income; or, to put the matter in a positive fashion, virtually everyone is assured a *minimum* income.

To explain adequately this movement would lead too far afield into social history, but some aspects of the matter may be mentioned. Urban industrial society as it developed here and in other capitalist countries produces its own peculiar hazards and insecurity for the urban dweller. Urban man is almost wholly dependent upon others for the production of the necessities of life, as well as for the income he receives to purchase these things. A money income affords him security in a restricted sense of the term. For the wage-earning element in an urban society the loss of income usually means impoverishment and distress in a short time unless external aid is made available. An assured and an adequate money income is therefore of vital importance to the urban dweller. However, there is no moral or legal obligation binding the employer to assure employment and income; employment depends largely upon the impersonal influences of the market. A change in consumer demand here, a new technological development there, a disruption or dislocation somewhere else, all or each may mean the loss of work and income to many persons. Loss of work and earnings for any other reason exposes the urban worker to hardship—industrial accidents, old age, illness, along with the hazard of industrial unemployment, are the principal causes of insecurity and hardship.

These hazards have of course always been present in any society, but urban life tends to aggravate their seriousness because the city dweller has but relatively restricted alternative opportunities to gain an income. He cannot as a rule raise his own food, and only a few are able or willing or have the opportunity to work for themselves. His efforts to secure an income and its attendant security are largely confined to the opportunities of working for someone else. The more able may find these opportunities sufficiently

abundant and certain, but many of the others do not. The actual or latent uncertainty confronting the typical urban wage and salary worker impels him to support measures that provide certain alternative sources of income —public work, social insurance, and public relief. These conditions stand in the background supporting the movement toward governmental assurance of income; they give substance to such political slogans as "income security," "job security," "the maintenance of work and purchasing power," and "full employment"—by government guarantee.

CAUSES OF INSECURITY

The principal causes that bring about loss of work and income are (1) unemployment, (2) industrial accidents, (3) illness, (4) old age, and (5) death of the family head. All except industrial unemployment contribute to a persistent problem of poverty and hardship; industrial unemployment, on the other hand, varies greatly depending upon the general level of economic activity. Even when unemployment is at low levels, the other factors create problems of poverty, but these are aggravated when unemployment is widespread. Some of the factors underlying loss of income will now be described.

Unemployment

Very broadly, two main components of the unemployment problem may be defined: (1) *the supply of labor or personal services available* for work; and (2) *the demand for these services,* including the demand of employers for wage and salary employees, and the demand of the "public" for the services of entrepreneurs and other self-employed persons such as doctors, lawyers, insurance salesmen, etc. Unemployment arises partly because of personal characteristics on the part of some forming the supply of labor and largely because of variations in the demand for labor.

Personal characteristics exert a selective influence on the labor market. Some persons have an aversion to sustained effort, others are physically or mentally unsuited to work, and some lack the capacity to get along with others. The incidence of unemployment is likely to be high among such persons, among whom are the "work-shy" and "semiemployable." These usually are the last to be hired and the first to be fired.

Variations in the demand for labor may be attributed to the following factors: seasonal changes, technological developments, business cycles, secular and structural changes, and frictions in the economy.

Seasonal Unemployment. Many productive processes are based on climatic conditions throughout a year. Agriculture requires a large labor force from spring planting to fall harvest; construction work normally lessens during freezing weather; manufacturing and trade may be subject to climatic influences, as well as custom and convention, e.g., the Christmas trade. Seasonal influences alone cause the demand for labor to vary normally by 3 or 4 million from the summer peak to the winter low. Seasonal unemployment for the most part is short-lived; some workers are able to reduce periods of idleness and income loss by shifting from one type of seasonal work to another. Some persons, like school youths, simply withdraw from the labor market until seasonal work resumes, and cannot therefore be classified as "unemployed." Nevertheless, seasonality of work accounts for many low incomes and the prevalence of a low standard of living even in periods of high levels of national income and employment.[1]

Technological Unemployment. Improvements in production and managerial techniques have long been recognized both as causes of labor displacement (usually in the short run) and, at the same time, of long-run growth in total production and in production *per man-hour* of work. As long ago as the eighteenth century, Adam Smith noted both effects; new machines "abridged labour" but also "freed labour" for other tasks and thus increased the productivity of labor. The "saving of labour" was equally evident to that stalwart champion of early nineteenth-century industrialism, David Ricardo, but he noted too the difficulties attendant upon the introduction of machines when he wrote that: ". . . the opinion entertained by the labouring class, that the employment of machinery is frequently detrimental to their interests, is not founded on prejudice and error, but is conformable to the correct principles of political economy." [2] English workmen at this time had frequently resorted to riots and machine-smashing raids in an effort to halt their displacement by machine methods. That new technology directly or indirectly holds the threat of unemployment *to individuals* cannot be questioned—too many skilled trades have been destroyed, too many industries wiped out, to admit of doubt.

Direct displacement may be illustrated by several examples. In 1920, all cigars were made by hand by skilled cigar makers; by 1933, more than half of the total output was made by machines operated by relatively unskilled

[1] Efforts to reduce seasonal unemployment are sometimes successful. In some cases seasonal work may be rationalized, i.e., spread out over the year, or additional lines of work or products may be added; but the major seasonal activities such as agriculture, construction, and other types of "outside" work must bow to the forces of nature.

[2] David Ricardo, *Principles of Political Economy and Taxation,* 3d ed. (London), 1821, p. 383.

workers. Because of technological changes in the telephone industry, 35 per cent more phone calls *per man-hour* could be handled in 1935 than in 1929. In the railroad industry the average work done per week by an employee increased 75 per cent from 1920 to 1936 because of improved technology.

Indirect displacement is no less far-reaching in effect. For example, a certain firm in an industry introduces new machines and equipment and is thereby able to produce its product at a substantially lower cost and sell it at a lower price. Its sales rise as a result, and it expands its employment, but some of its competitors may be driven to bankruptcy. If employment in the other firms is sharply cut or wiped out entirely, the introduction of new machines in one firm may be said to have indirectly displaced employment in others. A new product which creates a new industry and increased employment may also drive out another product or reduce its market, causing workers to be displaced (i.e., demand to fall) in an older industry.

The loss of work, the destruction of skills, the reduced income, and frequent hardships of those displaced (if only temporarily) are among the important costs of technological change. The other side of the ledger, however, shows the great increase in *man-hour* productivity, output, and employment over the years as a result of this change. From 1929 to 1941 one estimate shows a 34 per cent increase in the nation's output per man-hour of work, or 2.5 per cent per year compounded.[3] Individual gains have exceeded losses in this process of growth, but in the short run and for particular persons and occupational groups the loss of work and income is an ever-present threat.

Cyclical Unemployment. The growth of industrialism in this and in other capitalist countries has been characterized by rhythmical unstability, or business cycles. This has been the mode of capitalist development.[4] The recurrence of depression constitutes a potent threat to the security of wage earners (and others too); in some respects it is the most serious of the various causes of unemployment. During the upturn of the business cycle the aggregate demand for labor increases; during the downturn it declines. In the *minor* cycle (average period about 3½ years from peak to peak) the decline in demand for labor may amount to from 1 to 3 million, but the decline is arrested usually within a year and is soon restored in the subsequent cyclical upturn. *Major* cycles (embracing two or three minor cycles) cause a greater amount and duration of unemployment. From 1929 to 1932 factory employ-

[3] S. M. Livingston, *Markets After the War*, United States Department of Commerce, Washington, D. C., 1943.

[4] See Chap. 7 for an appraisal of cyclical change in the economy.

ment declined nearly 40 per cent, and did not recover its 1929 peak until the first years of the war in Europe. Recurring depressions constitute the greatest threat to the security of urban industrial workers; this fact led to the adoption of unemployment insurance systems abroad and in this country, as means of preventing hardship, and in part to reduce the severity of cyclical movements.

Secular Unemployment. The persistence of large-scale and prolonged unemployment may be described as "secular stagnation or retardation," arising from chronically low levels of investment. Aggregate demand for labor fell far short of the aggregate supply during the prewar decade; the 1933-37 cyclical recovery in this country failed to attain relatively full employment, as shown in Chapter 6. This period was characterized by (1) a growth in labor supply of some 500,000 workers annually, (2) an increase in output per worker, and (3) a failure of aggregate demand (particularly investment expenditure) to raise *real output* above the 1929 level. The influences that held investment expenditures at low ebb have been discussed at some length; [5] it need only be said here that when investment is thus depressed for long periods the "abnormally" large and prolonged unemployment may be termed *secular.*

Structural Unemployment. The migration of industry from one region to another, the depletion of mineral and forest resources in an area, the shift in demand from the products of one industry to those of another, and the loss of foreign markets are among the structural changes in an economy. Unemployment arises as a result of such movements and the maladjustments caused by them. Thus the migration of industry from New England (e.g., textile industry to the South and the shoe industry to the Midwest) caused at one time prolonged unemployment in many New England towns. The cutting over of timber land and the exhaustion of mineral deposits have left many "stranded" workers in the Great Lakes region and the western states. The decline in the cotton export trade of this country impoverished many southern rural communities. "Depressed areas," "decaying industries," and "stranded" workers are consequences of structural changes, and until workers migrate to new areas, or new industries arise in place of the old, prolonged unemployment is found in the areas adversely affected.

Frictional Unemployment. Along with other factors creating unemployment, many "frictions" develop in the economic system because of its complexity. Sporadic strikes and temporary shortages of transportation equipment may hold up production of essential materials, and this in turn

[5] See especially Chap. 7 for a review of the mature economy and expansive economy explanations of this phenomenon.

causes disruption of production in industries dependent on these materials. The temporary shortage of key workmen, a breakdown or change-over of machinery, will cause the layoff of many workmen. These and innumerable other frictions are generally more evident in a prosperity period when industry as a whole works on "tight" schedules (as in 1946-47); the slack in a depression usually minimizes them. The workers involved may "still have a job" but they will not be working at it.

Unemployment Levels

The combined influence of the causes or conditions described above determines the level of unemployment at any given time. The dominant influence, however, is exerted by the volume of investment: it is of crucial importance in determining the level of output and employment in normal times. When there is a secular or long-term tendency toward investment growth, the seasonal, technological, cyclical, and structural influences on unemployment are minimized. In an expanding economy, the seasonal worker stands a chance of supplementing his normal seasonal work with other employment; the technologically displaced worker is better able to shift to a new occupation; the stranded worker may find other areas or industries worth going to; and the workers periodically unemployed through depression influences find the duration and severity of this unemployment lightened by the early recovery of general business activity. When, on the other hand, the underlying investment opportunities are weakened or impeded, the demand for labor in general is depressed, and unemployment resulting from each of the specific causes is aggravated.

Yet even under the most favorable circumstances there is an *irreducible minimum* of unemployment. In a "full prosperity" period unemployment cannot be eliminated entirely. Frictions will cause some loss of work; seasonal influences operate in good times as in bad; here and there an industry or an area may be depressed; and some technologically displaced workers may not yet have found other work. During the war unemployment did not fall much below 1 million, and in a full peacetime prosperity, such as 1947, between 1.5 and 2.5 million persons were unemployed. When unemployment is at this minimum the economy is in a state of relatively full employment of its labor resources. In such a period hardship arising from loss of work and income does not vanish, but it recedes greatly in importance. The principal means of assuring a measure of income security and therefore of mitigating hardship is unemployment insurance, the main features of which will be discussed later in this chapter.

Industrial Accidents

Industrial accidents are a common occurrence in any industrial system. Estimates of the number of disabling accidents causing loss of work run as high as 8 million annually—or about 10 per cent of the number of wage and salary workers in nonagricultural employment.[6] These accidents may be classified as follows: fatal accidents, permanent total disability, permanent partial disability, and temporary disability.

These accidents involve two types of losses to individuals and families: (1) the loss in wages, and (2) the cost of medical attention and possibly funeral expenses. Losses may range from the cost of forfeiting a few days' pay and perhaps slight medical expenses, to the complete and permanent loss of income in cases of death and permanent total disability, with the accompanying funeral costs and life-long medical and dependency costs. To provide a measure of financial protection, state governments beginning in 1911 established workmen's compensation systems (described below). Similar provisions are found in many countries.

The Old-Age Problem

The loss of earning power as a result of old age is among the greatest causes of low income and destitution. Undoubtedly it has always been so, although only in recent decades has this problem emerged as a "responsibility of society." In earlier times, the care of the aged was almost exclusively the responsibility of family and friends; at present an increasing share of this responsibility is assumed by government. This change is the outgrowth of factors largely associated with the rise of urban industrialism.

Urbanization has accompanied the relative decline in agriculture as a means of securing an income and as a context in which to live. Over half of the population now resides in cities and less than one fourth on farms; a century ago these proportions were reversed. The customs and mores of predominantly agricultural communities usually assure the aged members of a family a relatively secure means of livelihood. The family assumes as a matter of course the responsibility for support of its aged; the means of providing this support are in good part derived from the farm itself, and the aged persons are usually able to contribute to this support by performing

[6] The Social Security Administration in its Annual Report for 1945 estimates that on any day 7 million workers are unable to work because of injury and illness, with a wage loss of 3 or 4 billion dollars annually. On any day injuries alone would be the cause of a relatively small fraction of this number of man-days lost.

useful tasks about the farm. There is, moreover, space to accommodate aged members; or if not, the family often can readily expand the farmhouse.

In urban areas the condition of the aged is less secure in many ways. As already noted, urban dwellers depend mainly upon money income, yet the urban environment, compared with the agricultural, is generally less conducive to employment of older persons.[7] The industrial discipline places a premium on speed and sustained effort and for this reason many industrial employers, and others as well, have established "hiring age" limits which discourage the employment of persons beyond fifty or fifty-five and over. Moreover, minimum wage laws and union wage rates tend to discourage the employment of workers who have passed their most productive years; i.e., age is a selective influence working against the older group.

The support of the aged by their families encounters more difficulties in urban than in agricultural areas. Urban families are smaller and have fewer working members and therefore fewer "hands" to support the older persons, who themselves are less useful in the urban environment. The financial demands on urban families, their normally crowded living quarters, and a weakening of family bonds, all contribute to the insecurity of the aged.

Moreover, changes in the composition of the population[8] have brought about an increase in the number of older persons in relation to the total population. One hundred years ago persons aged sixty-five and over comprised less than 5 per cent of the total; at present they are more than 10 per cent and are increasing. When this group is numerically and relatively small, its support is no great burden; it becomes a large financial cost as its size increases.

Other Dependent Groups

Problems of poverty and dependence arise because of ill-health, physical incapacity, the loss of the family income-earner who leaves dependent children, and other difficulties, both family and social. Again, such problems have always existed, but in the present era they increasingly command public attention; large governmental and private organizations have been created to provide support. Together with the unemployed, the injured, and the aged, these groups comprise a nonparticipating segment of the population whose support is derived in large part from public funds transferred by taxation from the incomes of participants.

[7] This point is supported by the fact that, in 1940, only 37 per cent of aged men in urban areas were in the labor force, in contrast to 60 per cent on farms.

[8] Discussed in Chap. 3.

PRINCIPLES UNDERLYING INCOME-PAYING POLICY

Once government embarks upon a policy of assuring some measure of income security, the underlying principles and methods of effecting such a policy must be clarified. Who should get government-supplied income? How is the amount of income to be determined? What means should be used to pay it? What are the purposes of the income-paying policy?

Bases of Eligibility

Several principles may be applied to determine who should receive income from government:

1. Only those persons or families in *need.*
2. Those who had worked in specified employments, to the exclusion of those in other employments.
3. Those who, while working, had earned less than a given minimum income.
4. All families and persons.

Each of these principles, or any others, must of course always be qualified in practice by the ultimate purpose of the government undertaking income assurance as this might be affected by prevailing public sentiment and political interests.

When *need* is the standard, the number receiving governmental income is likely to be kept to the minimum; the number increases progressively as the other standards are applied. In this country, need was the sole basis for the receipt of relief until the workmen's compensation laws were passed. Workmen's compensation programs adopted employment as the basis of eligibility, i.e., only persons employed in specified industries were eligible for its benefits. Unemployment compensation and old-age and survivors insurance under the Social Security Act also follow the employment criterion. In these social insurance programs, sizeable groups are excluded from benefit eligibility because they work in employments not covered by the legislation. At present, then, need is the basis used in relief programs, while the type of employment governs eligibility under social insurance. As for the remaining two principles, Britain and other countries base eligibility in part on the size of individual income, and Britain, mainly, has made some provisions toward universal coverage.

Determination of Income

Once the selection of the eligible groups has been decided, the question then arises: How is the amount of income to be determined? The following principles are prominent among those that may be adopted:

1. Equal income to all who are eligible.
2. Income based on need.
3. Income related in some way to previous income.

An *equal income* to all has as a principle the virtue of simplicity. Its most comprehensive advocacy has been the celebrated Beveridge Report proposed in the British Parliament in 1944. Objections to this principle of governmental income assurance are based partly on the logic supporting the other two principles and partly on practical administrative grounds. A flat allowance adequate for a single person would be inadequate for a family of six; and an allowance suitable to the needs of a family in rural areas would undoubtedly be inadequate for an urban family of the same size. Should a family previously earning $3,000 a year in private employment receive no more from government than another family that had managed to get by on $1,500? It can be argued plausibly that the income from government should be related to previous earnings. Other questions and objections arise as well. Can an equal income be accepted by public sentiment in a country where income differentials are great? To which racial group, occupation, and area would a flat income payment be related? The obvious problems involved in it have discouraged the extensive application of the principle of uniformity in this country.

The payment of income in accordance with *need* is almost universally accepted in programs of income assurance. Governmental income payments to families or persons vary directly with their needs, as modified by the size of a family, its composition (i.e., ages of children), the amount of its other incomes, its general health, and the living cost in its home locality. Determination of need—the amount required to relieve distress but no more—requires investigation; the traditional "needs test" is based on this fact. Determination is no simple matter, for it requires agreement on a minimum income and a plane of living regarded as essential or desirable. "Health and decency" are often regarded as constituting this minimum, but these terms are not readily definable, for medical authorities, nutritionists, and sociologists may frequently disagree, as may the public, on what they constitute. The principle of need, therefore, is a vague and controversial basis for the

assurance of income by government; it is a theoretically insoluble problem; nevertheless, in practice, a rough approximation of relative needs governs a large part of governmental income payments.

Relating governmental income payments to *previous income* is standard practice in the social insurance systems in this country. The inequality of income prevailing in normal employment is retained in those incomes received under social insurance. The skilled worker or the salaried man may receive two or three times as large a benefit as the low-paid unskilled worker. This fact is justified on the grounds that (1) the higher income groups are accustomed to a higher standard of living and therefore need more than the low-income group, and (2) the higher income groups pay more in taxes to the social insurance fund than do low-income groups. In the actual operation of social insurance systems this principle is modified so that the relative inequality of previous incomes is reduced. Modification is occasioned by considerations of need, since the total amount of money usually available is limited; if inequality was strictly maintained the lower income group would receive only a token income. The determination of income in *productive* employment is more rigorously defined than in the nonparticipating or nonproductive status. Limitations are imposed by productivity on personal services by the market relationships, while governmental income payments to nonparticipants in the productive system obviously have no such guide; instead public sentiment, conceptions of need, political pressure, and other considerations rule in the determination of this income.

Methods of Assuring Income

The means employed by government to assure a measure of income security likewise vary. The principal methods are relief, social insurance, and public employment. These represent not only different methods available to government but also different claims of the individual to governmental income payments.

A *relief payment* is a public gratuity. Need—destitution or the threat of destitution—is the claim of the person or family to this gratuity. This claim is not a legal right to relief payments established by law; public relief agencies merely have the obligation to relieve distress as they see it. The payment of relief involves no *quid pro quo;* nothing in return is expected of the recipient, and nothing is required of him in advance—destitution is the only prerequisite.[9] The number of persons receiving such payments, the

[9] There are, of course, legal requirements of residence in a county or state for a given period of time.

total amount paid out for this purpose, and the timing of this flow of governmental income depend upon the severity of destitution and the current attitude of government toward the problem.

The principal relief programs now in operation are: (1) *old-age assistance,* a co-operative federal-state relief program providing income to more than 2 million destitute aged persons; (2) *aid to dependent children, the blind, and other handicapped persons,* numbering some 600,000 persons or families and financed jointly by Federal and state governments; (3) *state and local general relief* (or poor relief), under which a half a million persons or families received aid during 1946. These programs are an outgrowth of the traditional local poor relief system which had its origin in colonial enactments, which in turn were patterned after the Elizabethan poor laws of 1601. Recent trends in the movement toward income security have been away from the traditional methods and conceptions of relief, partly because of the social objections to poor relief as demoralizing, and partly because of the broader economic purposes underlying recent income security programs.

Social insurance differs from relief in that (1) the social insurance recipient has a right to benefits based on law; (2) he receives the benefit automatically upon satisfaction of certain legal and administrative conditions; (3) the benefit is paid whether he needs it or not; (4) the amount of the benefit is related to previous earnings; and (5) previous contributions must be made by him or his employer, or both, to the social insurance fund from which benefits are eventually paid. Poverty is not a requisite; the well-to-do are as eligible as the poor. No social stigma is attached to these payments because the recipient has a "right" to them, and in some cases, he has paid taxes for the financing of these programs.

These features render social insurance more desirable than poor relief on *social grounds.* Under poor relief the recipient must first exhaust his personal resources—savings, personal property, and private charity—before he is eligible to receive relief payments. Social insurance programs usually avoid this painful and humiliating process; indeed, the prevention of destitution is their principal purpose.

Public work is a method of income assurance quite different from the others; it involves a direct *quid pro quo*—wages are paid in return for productive service. Two types of public work programs may be distinguished: (1) *work relief,* and (2) *normal public works.* Work relief featured prominently in the prewar decade, when as many as 4 million persons were employed at one time on work relief projects. The purpose of such a pro-

gram is to give employment per se; the work accomplished is a secondary consideration. Normal public works, on the other hand, lay the emphasis on projects, or the completed work per se, and not on rendering direct employment. However, indirect employment (in producing materials and equipment for public works projects) may be substantial. Another contrast may be noted. The basis of employment on work relief is need; on public works this is no consideration, for work is offered under the hiring conditions normally prevailing in the construction industry.

Purposes of Income Assurance

The purposes of the assurance of income have been suggested in part by the foregoing discussion. These purposes include:

1. The alleviation of distress.
2. The prevention of distress by maintaining given standards of consumption.
3. The maintenance of total consumer expenditure; this involves an increase in the spending ratio (a decrease in the saving ratio) for the economy as a whole, in order to contribute toward (a) increased economic stability, and (b) increased levels of output, income, and employment.

The *alleviation of distress* was the sole object of the traditional relief system here and elsewhere, and it remains so in the present programs of relief. The growth of social insurance, however, marked the appearance of a policy for the *prevention* of distress or destitution. To accomplish this purpose it is necessary that (1) the population be adequately covered by social insurance, (2) all hazards that lead to destitution be encompassed in it, (3) benefit payments be made promptly, and (4) these benefits be large enough to cover minimum family needs.

The maintenance of *total consumer expenditure,* or purchasing power, attained prominence in this country during the 1930's, especially as part of the mature economy proposals.[10] In this thesis, relief, social insurance, and public work programs are regarded as means of bolstering consumer spending, provided that their funds are derived (1) in short periods from governmental borrowing, and (2) over long periods from taxes on high incomes. This involves basically a redistribution of income in favor of the lower end of the income scale, and a consequent increase in the spending ratio for

[10] See Chap. 7.

the economy as a whole.[11] The purpose involved here is economic and is predicated upon the view that economic expansion in this country requires a diminution of the savings ratio in favor of consumption. This view, however, is debatable. If expansion maintains its former stride—with no governmentally induced increase in the spending ratio—then this economic purpose of security programs loses its *raison d'être* as a national policy measure. And if pursued under these circumstances, it would perforce retard the growth of capital and long-run expansion. This is precisely the fear of those who dispute the mature economy view; for them the assurance of income should be limited to social purposes only.

It should be observed that economic expansion in other countries, particularly those that suffered extensive war damage, may require a *reduction* in their traditional consumer spending ratio. Under these circumstances, a growth in social insurance and related programs might impede long-run economic expansion. For example, the increase in Britain's social security program is likely to be at the expense of capital growth and long-term expansion of output. The rebuilding of Germany, too, would be retarded by a substantial growth in social insurance and relief.

THE PRINCIPAL PROGRAMS

The discussion turns now to a brief description of the main programs of income assurance. First to be considered are those established for the protection of wage earners: workmen's compensation and unemployment insurance. This is followed by a description of old-age insurance, survivors' insurance, old-age relief, and other programs.

Workmen's Compensation

Accident insurance was the first form of social insurance to be adopted in this country. It was established gradually by state legislatures because of the inadequacy of the traditional employers' liability laws in meeting the social problems arising from industrial accidents. Consider briefly the main features of the employers' liability laws: (1) they sought to determine who was responsible for causing or contributing to accidents; and (2) payment for injuries sustained was determined by litigation. The injured workman was forced to institute damage proceedings against the employer, if no

[11] As mentioned in the preceding chapter, this objective is sought not only through governmental income-paying programs but also through union wage policies, minimum-wage legislation, and related social and economic legislation.

voluntary and acceptable settlement could be made. In any given case the court might fix responsibility for the accident on (1) the employer or his agent (foreman or supervisor of the injured), (2) the injured workman himself, (3) a fellow worker, or (4), in its wisdom the court might find the accident to be "an act of God" and therefore beyond human responsibility. The injured person obtained damages only if the employer or his agent was found by the court to be the responsible party. Litigation was usually a lengthy and costly procedure; meanwhile the injured person and his family often had to resort to private charity or local poor relief. And often, too, the injured person received no damages.

The purpose of workmen's compensation legislation is to prevent distress by providing a prompt and assured payment to the injured worker or his family without raising the question of legal responsibility for the accident.[12] This legislation endeavors to make industrial accidents a cost of production, paid directly by industry. Employers are required to insure against injury of employees engaged in actual work; premiums are paid by employers either to private insurance companies or to state insurance systems which pay medical expenses and compensation benefits to injured workers or their survivors. These premiums, which are similar in all respects to the cost of insuring against fire, theft, or other risks, are related directly to the risk of injury; the greater (or lesser) the incidence of disabling accidents the greater (or lesser) is the cost of workmen's compensation premiums to the employer. Legislation thus has provided a financial incentive to industry to reduce accidents through safety and accident prevention measures.[13]

From the employer's point of view, workmen's compensation removes the question of responsibility and the problem of damage suits.[14] The risk, as measured by his premium payments, becomes a certain or definite cost of his business operations which he seeks to cover in the sale price of his products. The injured worker, on his part, has the assurance of an income, in lieu of a possibly larger but more uncertain and delayed damage payment from his employer under employer's liability legislation. Workmen's compensation is not a full restitution of the wage losses sustained by the injured worker; this is justified on the grounds that his own responsibility (if any)

[12] This point needs qualification. State laws usually provide that in cases of gross or willful negligence on the part of the injured no compensation is payable.

[13] The incentive feature in this form of social legislation and the consequent reduction in accident risk had a profound influence on the formulation of unemployment compensation legislation in the 1930's, as is indicated on pp. 883-84.

[14] Except in employments excluded from workmen's compensation and in special cases covered by compensation, e.g., if willful negligence is involved.

is usually disregarded and he is not required to show proof of employer responsibility.

Most of the wage earners in manufacturing, mining, trade, service, and construction industries are covered by workmen's compensation. In case of injury while at work, weekly benefit payments are made to the injured worker ranging from one half to two thirds of his weekly wage. Such weekly benefits, however, are subject to maximum and minimum limits; the maximum in some states is as high as $25, the minimum as low as $5. If the accident causes permanent total disability a few states provide benefit payments for life; usually, however, payments stop at the end of five or seven years. Permanent partial disability (e.g., the loss of an arm) is compensated normally on the basis of a schedule of payments for prescribed types of injury. When an accident is fatal the survivors are usually paid weekly benefits for five years and sometimes longer. Medical and hospital expenses are provided in the state laws, ranging from a maximum of $100 in some states, to "all necessary aid" in about one third of them.

Workmen's compensation, of course, is not intended as a means of compensating total losses and costs; indeed in only a few states is it likely to cover as much as half the wage losses and other costs, and in some of them the percentage is much lower. Its purpose is to offset enough of the losses to prevent the pauperization of the family—enough to tide the family over until the injured person is able to return to work. Before this purpose is attained for workers covered by insurance, however, much reform is required in half or more of the states; and all states should extend coverage to groups now excluded if this purpose is to prevail for all who are subject to work accidents. All state systems, too, generally ignore occupational disease as a cause of disability, and this is a grave shortcoming.

Unemployment Compensation

Unemployment insurance or compensation was established in this country as an outgrowth of the unemployment crisis of the 1930's. Before this time, many reformers had advocated this means of income assurance, patterned after the plans in effect in Great Britain and elsewhere. However, employer opposition, trade-union opposition and indifference, and the lack of general public support prevented its adoption. The Social Security Act of 1935 encouraged the adoption of unemployment compensation by all the states.[15]

This act prompted the adoption of unemployment compensation by the

[15] Except the Wisconsin system, which had already been established by the state in 1934.

states in the following way: a federal excise of 3 per cent is levied on the pay rolls of employers with the proviso that they may deduct—up to 90 per cent of the amount of this tax—any state tax they pay for unemployment compensation. Suppose the federal pay-roll tax on an employer is $1,000, and that he also pays $900 to his state government for unemployment compensation. The law permits him to credit this latter amount against the federal tax, reducing it to $100. If a state had no unemployment compensation system, the entire $1,000 would be paid to the Federal government for possible disbursement to the unemployed of not only his but all other states. To avoid this and to relieve its unemployed, each state promptly and "voluntarily" established an unemployment compensation system.

In addition to the tax credit or offset, each state receives a grant from the Social Security Administration for the administrative expenses of its unemployment compensation system provided it meets certain federal requirements. These include (1) suitable administrative methods, (2) payment of benefits through federally approved public employment offices or other public agencies, and (3) the transfer of state unemployment compensation tax funds to the safekeeping of the Unemployment Compensation Trust Fund in the Federal Treasury.

The federal excise on pay rolls is levied on those employers who employ 8 or more workers for at least 20 weeks during a year. The tax rests on each employee's wage or salary up to $3,000 annually. All employment, however, is not subject to this federal tax. In addition to employers hiring fewer than 8 workers and employing workers less than 20 weeks during a year (if only for one day a week), the law exempts such employment as: maritime labor; service in nonprofit religious, charitable, and educational institutions; domestic service in private homes; agricultural labor; all government employment; and railroad workers (who have their own unemployment compensation system). As a consequence some unemployment is not comsated at all, principally among the lowest paid wage-earning groups in the economy—agricultural workers, domestic servants, many employees working in small establishments, and workers who depend upon casual employment.

The benefit provisions of the state systems may be summarized. Benefit payments are approximately half the average weekly earnings of the unemployed worker in "covered employment," but each state establishes a maximum limit which ranges from $15 in some states to $28 in others. Approximately three quarters of all employees covered by these systems in 1946 resided in states with maximum benefits of at least $20. The maximum

number of weekly benefit payments that may be made during a year to an unemployed worker ranges from 14 weeks in some states to 26 in others. Actual benefit payments now average $18 to $20 weekly.

Financing Unemployment Compensation. Unemployment compensation is financed wholly by specific taxes or contributions imposed on employers for this purpose by the states.[16] In all but one state this tax is 2.7 per cent of the wages paid by enterprise. Since this is 90 per cent of the 3 per cent federal excise on pay rolls, employers are able to credit their entire state tax against the federal tax. The funds thus raised by the states are deposited in the Unemployment Compensation Trust Fund to the account of each state and invested in federal interest-bearing securities. The money in this fund constitutes the *reserves* available for the payment of unemployment compensation; in fact, the fund is comprised of 51 reserves for the 48 states, Alaska, the District of Columbia, and Hawaii. As money is needed for the payment of benefits, the Federal government transfers it to the states, charging each state's reserve for the amount transferred. Tax receipts, therefore, add to the reserves of the states held in the fund, and benefit payments by the states draw their reserves down.

Although the normal state tax is 2.7 per cent of pay rolls, most of the states permit a reduction in this tax for employers who maintain stable levels of employment. This is termed *experience rating* or *merit rating*. Thus, when the taxes paid by an employer exceed the unemployment benefits paid to his discharged workers, his tax rate is reduced. Just as the prospect of a lower premium induces industry to minimize the hazard of industrial accidents, tax reduction under experience rating endeavors to encourage industry to reduce the hazard of unemployment. The parallel between industrial accidents and unemployment, however, is at best a limited one. The reduction of accidents is subject to control by the management and workers in a plant through safety measures, discipline, and education. But the causes of unemployment are largely beyond control of the management and workers of a plant; these causes are governed primarily by the market for the goods produced, i.e., by changes and shifts in demand for the products of industry, and by other causes described above. The principal effect of experience rating is to reduce the unemployment compensation tax rate on industries that are inherently stable, while it can do little toward stabilizing employment in industries subject to large seasonal and cyclical movements. It should be noted that "stabilizing" employment is not synony-

[16] In a few states employees are also taxed, but this accounts for only a small part of the total funds raised for unemployment compensation.

mous with reducing unemployment, although experience rating assumes these to be virtually the same thing.

Tax receipts flowing into the reserves vary in accordance with the level of employment and pay rolls; the flow increases in a period of prosperity and diminishes in a period of contracting employment and pay rolls. The withdrawal of funds follows an opposite course, decreasing in a prosperous period when unemployment is low, and increasing in a recession period as unemployment mounts. The tax and benefit provisions of an unemployment compensation system should be designed to build up a reserve in good times for disbursement in periods of need. From their inception in the mid-1930's through the war years, the state systems in this country have had reserves adequate to meet their needs. In fact, total reserves aggregated 7 billion dollars in 1947.

Unemployment compensation is frequently referred to as the "first line of defense" against unemployment. The expression is apt, for the eligible unemployed receive benefits within a few weeks after loss of work, but exhaust their benefit rights usually within four to six months. Benefit payments during this period prevent undue hardship, and if a slump in employment is of short duration, the program is effective in tiding over the period of idleness. However, unless private or public work is available at the expiration of the benefit period, the unemployed face the prospect of destitution. Unemployment compensation is of no help to the long-duration unemployed.

By what means can income be made available to workers whose benefits have been exhausted? One obvious step is to increase the length of the benefit period, as some states have done, but there arises the difficulty of maintaining payments from a fund supplied by a fixed rate of taxation. If additional money were made available by appropriations of Congress, payments could be extended greatly; this would of course convert unemployment compensation into a virtual relief program, as was done in Great Britain in the late 1920's and early 1930's. Relief and public employment are alternative means of governmental assurance of income to those who have exhausted their benefits and to those who are by law excluded from unemployment compensation.

Unemployment compensation not only mitigates the hardships of short-term unemployment but it also acts to cushion the decline in purchasing power at the outset of a depression. Nearly half of the wage losses are offset by benefit payments, thereby lessening the deflationary impact of unemployment on consumer expenditures and on production and employment in the consumer goods industries. In a minor recession this influence may

moderate the decline and hasten revival, but it cannot maintain consumption during a major depression.

Old-Age and Survivors' Insurance

Old-age insurance was inaugurated by the Social Security Act of 1935; extensive amendments to this federal statute in 1939 modified this program and added insurance benefits for the survivors of those covered by the act. There follows a brief description of this comprehensive system of social insurance.

The Old-Age and Survivors' Insurance system is a federally operated program nationwide in scope which covers the bulk of the wage and salary workers of the country. Employment in manufacturing, retail and wholesale trade, privately owned public utilities (except railroad transportation), mining, the service trades such as service stations, hotels, beauty shops, and theaters, come under the legislation. Large occupational groups are, however, excluded, such as: farm labor, domestic servants, public employees, railroad workers, employees in nonprofit organizations such as educational, charitable, and religious institutions, and self-employed persons such as farmers, doctors, lawyers.

Persons covered by the old-age insurance system receive retirement benefits or annuities at age sixty-five if they retire, or later upon retirement. The amount of the benefit depends upon previously earned income in covered employment and the number of years a person has worked. The Social Security Administration keeps records of wages and salaries earned by persons covered by the program; from these records it determines each person's average monthly income, and upon retirement the benefit is calculated as follows: (1) 40 per cent is taken of $50 of this average income, and (2) 10 per cent is taken of the excess over $50 and up to $250. The sum of these two calculations is then increased 1 per cent for each year the insured person has been covered by the program; this is the monthly annuity (primary benefit) for a single person. The aged wife of an annuitant is given a supplemental benefit (wife's benefit) equal to half the primary benefit of her husband. Upon the death of the annuitant, the widow continues to receive a benefit until her death.

When a person in this program dies before his retirement, provision is made for survivors' benefits to his widow and dependent children, and in some cases to his dependent parents. Benefits to the surviving family cannot exceed $85 monthly or fall below $10. Table 94 gives a summary of beneficiaries in 1946.

Table 94 **NUMBER AND TYPES OF BENEFITS AND AVERAGE BENEFITS: OLD-AGE AND SURVIVORS' INSURANCE, June, 1946**

Type	Number	Average benefit
Individual old-age annuitants	427,000	$23.80
Individual and wife	193,000	38.80
Individual and one or more children	11,000	40.90
Widow only	115,000	20.20
Widowed mother and one child	66,000	34.20
Widowed mother, two or more children	57,000	50.00
Children only	113,000	22.90
One or more parents	6,000	14.20

Source: Social Security Administration.

The Financial Arrangements

This social insurance system is a compulsory, contributory system which is financed by taxes levied on persons in covered employment and on their employers. Each person is subject to an income tax amounting to 1 per cent on his income up to $3,000 a year; income in excess of $3,000 is not taxed. Each employer, in turn, is taxed 1 per cent of his payroll, but not on any individual wage or salary in excess of $3,000 per year. The system, therefore, raises half of its funds from those who eventually will receive benefits and half from their employers. As in the other forms of social insurance already discussed, the employer tax is added to the cost of doing business; and the tax on individuals may be viewed as "compulsory savings" on their part for the purpose of old-age security.

Tax receipts go into the Federal Treasury but are not "earmarked" or set aside for old-age benefit purposes; they are spent as promptly as government spends its other tax receipts. The old-age insurance system obtains its money, then, from an annual appropriation from Congress in an amount approximately as large as the old-age tax receipts. The money thus appropriated is used to pay benefits and cover the other costs of the system; excess funds go into the Old-Age and Survivors' Insurance Trust Fund in the Federal Treasury and are invested in interest-bearing United States securities. The interest payments on these securities accrue to the fund.

From 1937 to the latter part of 1946, appropriations to the fund had

totaled 8.6 billion dollars, and the interest earned on fund securities amounted to 600 million dollars. This total of 9.2 billion dollars represents the money available to the old-age system during the decade following its inception. Expenditures from this fund for benefits and costs of operation amounted to 1.3 billion dollars for the same period, leaving a reserve in the fund of nearly 8 billion dollars.

The cost of the Old-Age and Survivors' Insurance system will in the course of time amount to several, and possibly many, billions of dollars a year. For this reason the financial arrangements are important. Several major forms of financial provision are possible. Two of these deserve brief comment, for they portray the financial issue at hand. One may be termed *full reserve* and the other, *pay-as-you-go.*

The full reserve plan is one that takes into account the future liabilities (i.e., future demands for benefits) of the system and builds up a reserve to meet them. Each year the system operates liabilities grow larger as people at work increase their benefit rights. The full reserve plan would add each year an amount to its reserve sufficient to cover in full future claims. A full reserve plan, therefore, requires large annual revenue (presumably by taxation) and heavy investment in securities to accrue interest. In the original act, the old-age insurance system embodied these principles, but opposition forced the full reserve to be abandoned in 1939.

At the other extreme, the *pay-as-you-go* plan makes no reserve provision for liabilities currently undertaken which must be met in future years. This plan merely raises each year sufficient funds to meet current benefit claims.

The issue involved between these plans is intensified by the fact that Old-Age and Survivors' Insurance is a system which begins its operations on a small scale and later, after two or three decades, must pay out great sums in benefits each year. The pay-as-you-go plan would require higher social-insurance tax rates in future years than at present, while the reserve plan, by building up the fund with direct appropriation and interest payments, would seek to keep taxes relatively equal between early and later periods.

The present Old-Age and Survivors' Insurance system falls somewhere between these opposing plans. Its reserve is termed a "contingency" reserve, not a full reserve; its tax rate is higher than required for current operations, but not high enough to create a full reserve. Therefore the present tax rate, plus interest on its contingency reserve in the future, will undoubtedly provide less money than will be needed to pay benefits in the future. At some future date social-insurance tax rates either will be increased, or the added revenue will be obtained from other tax sources or from borrowing.

Other Old-Age Insurance Plans

Railroad employees have a separate old-age retirement and survivors insurance system, as well as a separate unemployment insurance plan. This system is federally operated and covers all railroad and related workers, numbering at present nearly 3 million.[17] It provides retirement and survivors insurance paralleling the social security system, and in addition grants disability benefits. To finance the plan the Federal government levies a 5¾ per cent tax on the wages and salaries of covered employees and the same rate on the pay rolls of the railroads. In both cases, individual income in excess of $300 monthly is not taxed. The tax rate on both employees and employers is much higher than that of the social security system—5¾ per cent in contrast to 1 per cent—but in return for the higher tax railroad employees receive more liberal benefits. In 1946, 180,000 persons received an average of $104 monthly under this retirement plan.

Other social insurance systems such as the retirement plan of the Civil Service Commission for federal employees cover special groups of retired and disabled persons. This is also a compulsory, contributory system; employees and government both contribute 5 per cent of salaries and pay roll, respectively. In 1946 about 10,000 retired federal employees received an average monthly benefit of $80. In addition, many state and local governments have established retirement plans for their employees. Retirement, disability, and survivor benefits under the Veterans Administration are paid on a large scale; in 1946 the number of beneficiaries exceeded 3 million, at a total cost above 1.5 billion dollars.

While the retirement and survivor systems described thus far are governmental programs, a large number of private retirement plans have been established by employers which are often financed by employers and employees jointly. Most of these plans have been initiated by corporations, although in some cases under the impetus of union pressure on employers. Many private plans have either been initiated or adapted to supplement the somewhat inadequate benefits provided by the Old-Age and Survivors' Insurance system.

[17] Employees of railroads, sleeping-car, and express companies subject to the Interstate Commerce Act, and of companies controlled or owned by the carriers and engaged in performing service in connection with railroad transportation, are covered. Employees of railroad associations, national railway labor organizations, and employee representatives are also covered.

Governmental Relief Programs

In addition to old-age insurance, the Federal-State Old-Age Assistance program provides relief payments to persons aged sixty-five and over who are in need. This program is operated by the states under standards established by the Federal government and with the financial assistance of the Federal government. It is a noncontributory program and its benefits are not contingent upon or determined by previous employment. It is a complementary measure to old-age insurance, for those excluded under the latter system may receive old-age assistance payments; and old-age annuitants who receive inadequate insurance benefits may have them supplemented by this relief program.

This relief program is an outgrowth of the old-age pension plans that had been in operation in some states before the depression of the 1930's. Federal grants began with the Social Security Act of 1935. In the latter part of 1946, 2.2 million persons were aided; the average old-age relief payment to this group was approximately $30 per person per month. These payments are not received as a matter of legally established "right"; government merely assumes the responsibility of alleviating distress.

State programs, aided by federal grants, have been established to provide incomes to blind persons and dependent children in need. In recent years the number of dependent children whose mothers have received public money for their support has varied from 600,000 to 1,000,000 a year. In addition, state and local governments give relief payments to several hundred thousand persons in need who fail to qualify under any of the other programs described previously: those not old enough to receive old-age benefits, those ineligible to receive unemployment insurance benefits, those who have exhausted their benefit rights—the poor who are unable to work and earn an income but have no private means of support.

SOME ECONOMIC CONSIDERATIONS

Social insurance and related programs involve both a *money cost* and a *real cost* to individuals. Most of the money costs are specific taxes, such as the social security and railroad retirement taxes. Others are not specific, such as the costs of old-age relief and aid to dependent children which are financed by general tax revenues. The costs are therefore compulsory on those taxed, and currently amount to several billion dollars annually.

The money thus raised involves a transfer of income from persons taxed

to those receiving benefits; the latter are able thereby to consume more, while those who bear the tax spend (or save) less. Those who *bear* the tax may not be those who *pay* it in all cases. The tax on employers for social insurance may be borne by the owners of industry, or passed on to the buyers of their products in higher prices, or offset by increased economies in production. In any case, however, governmental assurance of income to nonparticipants involves a transfer of real and money income from those who bear the tax to those who receive benefit payments.

This point bears directly on the effect of social insurance on purchasing power. The opinion is frequently expressed that social insurance "creates" purchasing power. This is an illusion so long as the benefits are financed by taxation and other contributions. There is no creation, but merely a transfer of income and consuming power from one person to another or from one group to another. The effect of the transfer at any given time may encourage consumption in the economy and reduce savings. But this might also reduce consumption in the long run. This is not to say that a reduction in savings because of a transfer of income *necessarily* means a lessened consumption in the future. Under some circumstances it may mean more future consumption; i.e., if current investment opportunities are meager and economic resources are not fully used, a transfer of income from savings to consumption expenditure would tend to maintain a higher aggregate demand and therefore output, employment, and income. Under these circumstances, present consumption will be—and future consumption may be—increased by the transfer of income. On the other hand, if current investment opportunities are abundant and resources are fully employed, a transfer of income from savings to consumption expenditure would lessen the funds and the economic resources available for investment purposes. This would restrict the future output of consumer goods and therefore consumption. As already noted, large social insurance commitments in Britain threaten to retard future capital expansion in that country.

The distribution, or sharing, of costs among economic groups is a matter that will grow in importance in future years because aggregate costs will increase enormously. Two questions arise: Who *should* pay the costs (taxes)? and Who *will* bear the costs? The answer to the first question is in practice largely a matter of political expediency, and the answer to the second is given in part by the shifting and incidence of taxation.[18] As already shown, employers in this country pay the bulk of social insurance taxes and contributions, while employees pay a relatively small part. The employee is taxed because he and his family will receive benefits; in effect, such a tax

[18] See Chap. 33.

asserts the individual's responsibility to save and provide for himself, but compulsion is exercised to assure that he does. The employer tax rests partly on political grounds, and partly on the assumption that low-income groups cannot pay in taxes as much as is needed for social insurance. Implicit also in this division of taxes is an ethical premise that those more able should help those who are less able. The employer tax, however, is probably shifted to some extent. That part which can be and is shifted may appear in the price of commodities and services in the long run, although the shifting may take other forms.

In some other countries the cost of social insurance rests more heavily on employees than is the case in this country. However, in other countries the national governments usually assume part of the cost from general tax revenues. This practice has arisen largely because the burden was believed to be too great to be borne by employees and employers alone through specific taxes. To add to employer costs might, by increasing production costs, reduce the ability of producers to sell their products in foreign markets. To add to the employee tax might create political difficulties and unrest. As social insurance develops and expands its costs grow also, and the problem of sharing the cost assumes political and economic importance.

As already noted, one basic economic consideration is the influence of social insurance taxes and expenditures on spending-saving ratios. Another is the influence of the method of distributing costs on economic incentives. The tendency in this country is to shift some part, and possibly a large part, of the cost of social reform to the high-income groups by means of the progressive income tax. This tendency finds popular support among the masses of low-income recipients and those who speak for them. As social insurance costs increase, this view may grow in popular appeal, for a tax on a low or moderate income is no more popular to those who bear it than a high tax on high incomes. Politically the difference lies in the greater number of low and moderate incomes and the relatively small number of high incomes. Yet political expedience may lead to tax policies inimical to investment expenditure.

In an earlier chapter it was said that the controversy over long-run investment prospects was enlivened by strong sentiments relating to the strength of the underlying forces of expansion in the economy; the controversy has been noted more for the heat than for the light it has thrown on the subject.[19] As the social insurance program has developed thus far, its financing has probably had little effect on the spending-saving ratio

[19] See Chap. 7 for the several views on the question.

and upon long-run investment prospects. Prewar tax-reform programs and the necessities of war finance have had more influence on incentives and on spending-saving activities than has social insurance. As social insurance grows in importance, however, these economic considerations will grow also. In the long-run formulation of policy relating to the sharing of costs the following point should be stressed: social insurance and other transfer payments do not create purchasing power; they represent a shift in purchasing power from one group to another. Nor do these payments represent the production of any part of the national output; they are nonparticipating incomes. The material well-being of those who receive these benefits and the well-being of the rest of the population rests largely upon current production of consumer commodities and services and upon the production of capital facilities for future consumption purposes. As England has discovered, social insurance, however great its social advantages, is largely an unproductive enterprise. Since its success can be assured only by high levels of output and economic expansion in the long period, social security is necessarily subordinate to long-term economic growth.

PART IX

TYPES OF ECONOMIC SYSTEMS

CHAPTER 41

Proposals for Reform

This book has mainly been concerned with describing and analyzing the operation of the economic system of the United States, which has been variously referred to as a "free enterprise," a "price," or "competitive," or "capitalist" economy. At numerous points in the preceding chapters, particularly in "Prospects of Economic Progress" (7), "Price Control" (23), and "Government and the Economic System" (30), mention was made of other economic systems—either extant or theoretical. It has, for example, been pertinent to compare and contrast the methods as well as objectives of collectivist economies with those of capitalist economies in such economic processes as regulation of prices, allocation of resources, and public regulation of production. In the chapter on "The Prospects of Economic Progress," in which certain major points of view toward the adequacy of present-day capitalist system as a means of production and distribution and as a kind of social organization were described, the wide range of possible modifications and *alternatives* to this system is suggested. It is now necessary to examine the most significant alternate economic systems that exist in the modern period both in practice and in theory.

For most purposes, capitalism can be defined as an economic system in which: (1) production is carried on for a market, and each producer seeks to maximize profits; (2) resources ("capital") are privately owned and are freely disposable by their owners (free enterprise and freedom of contract); (3) some persons employ the services of others. The third point in the definition is often expressed as the "separation of the workers from the means

of production." Now, capitalism in this sense does not exist *throughout* the American economy, although it is of course the dominant institution. A considerable volume of production is carried out by government, which does not produce for a market, although even here there exists a separation of "the workers" from the ownership of "the means of production." Furthermore, this definition does not fit much of American agriculture, for although farmers produce for a market the owner-cultivator is both a laborer and the possessor of his "means of production." Nor does it apply to the independent professional people and to one-man business enterprises. It is also true that in its present form the American economy tolerates considerable restriction of freedom of enterprise and freedom of contract; public utility and labor legislation exemplifies such restriction. Moreover, the existence of labor unions has significantly changed the position of the wage earner vis-à-vis the employing owner of capital.

Thus the present American economy consists of a complex set of changing institutions and processes and laws; it perhaps cannot be labeled directly as "capitalism" in the terms of the definition here. However, several decades ago this definition would probably have sufficed with slight modification. Since it is a basic description—keeping in mind important qualifications—one can reasonably and with convenient simplification use it in comparing capitalism with other economies.

It is to be emphasized that capitalism is a system, an economic and a social system. Although there is no "economic constitution" which sets forth the framework of capitalism, nor a general code for its operation, capitalism is an organized method by which wants are satisfied within a society; it is recognized by the habits and, to some extent, morals of society and it finds political sanction *as a system* in various laws which either protect or assert (or even abridge) certain individual measures by which it operates.[1] Thus it is perfectly logical to speak of capitalism generally and to regard it as a concerted and defined economic organization of society.

SOCIALISM

Socialism is a proposal for an alternative economic and social system. It may be defined as a system in which all or a major part of economic resources are directly controlled by government ("government ownership of the means of production"), for the purpose of satisfying wants in ways determined by government ("planned production"). This definition is broad enough to embrace all of the important and most of the lesser of the many

[1] See Chap. 30, pp. 647-51.

schools of socialism. Socialists themselves have always emphasized that capitalism is a system, that the misfortunes, ills, and miseries of men are due to the system as such. It is the belief of the socialist that it is not within themselves but within the economic system that men's inhumanity to their fellow men has been found. It follows in the socialist view therefore that the system should be changed. Once this is done, and since there is never any question that the new system will not be better, the natural goodness of human nature will unfold itself and all major problems will be solved. The essence of the socialist position lies in three beliefs: the system is at fault, a better system can be created, a better system should be created.

Socialism is of course not the only proposal for reform. Many reformers find some one factor or other to be the main cause of human difficulties, factors such as alcohol, inadequate diets, insecurity, unequal incomes, etc., for the list is very long. Such reformers generally try to work for their goals within the framework of existing institutions. The present concern is not with them, but with those reformers who have proposed a radical transformation of the present economic system.

The Origins of Socialism

Modern socialism as a swirling mass of ideas originated in western Europe in the early part of the nineteenth century. Earlier in the history of men's thoughts there had been expositions of perfect societies, such as Plato's *Republic* and Sir Thomas More's *Utopia,* but these did not—important otherwise though they were—develop into political and social movements of consequence. The socialism of the nineteenth century was bred as a revolt against the capitalist society of that day and it became a political movement, growing with the very capitalism it criticized.

Socialism, like capitalism, has philosophical as well as political implications beyond purely economic considerations, but these can only be mentioned very briefly. The concern here is with the aspects of socialism as an economic system. Socialism is a secular system of thought, and in part its development may be attributed to the decline of the power of religion in the modern age. Socialist thought and antireligious sentiments were linked together in the nineteenth century; it is significant that some schools of socialism are based on a solely materialist outlook. The impact of the French Revolution (and to a lesser extent, the American Revolution) on the European mind must also be given some share in the formation of socialist ideas. The French Revolution smashed a political and social system which

had seemed strong and inviolable: it had gone beyond the hopes of even the revolutionaries. But in the end it proved disappointing, for although the autocratic power of the French Catholic Church and nobility had been broken by the revolt, the millennium had patently not arrived. The incipient socialist belief was that another revolution, an economic one, would follow; like the political overturn of the aristocracy in the eighteenth century, a revolt would depose the power of the wealthy in the nineteenth century.

The development of an *industrial* capitalism, particularly in England, allowed socialism to attack with great effect and at least humanitarian justification the living conditions imposed by industrialism in the early nineteenth century. The new industrial towns were crowded with people who had been forced out of their rural occupations and way of life by the great agricultural changes of the eighteenth century.[2] The new factories equipped with power machinery became a main cause of employment and stood as the symbol of capitalism. Profits were high, wages were low, and working conditions were—from the modern standpoint—unbelievably bad. Poverty and misery were prevalent for decades in the very period in which modern methods of production were being developed.

Capitalism established, moreover, a new system of relations among men, the relationship of contract and of individualism, as contrasted to the status of persons as part of a social class. Thomas Carlyle called it—with contempt—the cash nexus. Each individual in a capitalist society was to act for himself; no one was responsible for another. The employer was not responsible for his employees, as the landowner had once been for his retainers on the land. Capitalism of the early nineteenth century came to be regarded by many as a harsh and inhuman system.

Some reforms were accomplished, as in the several reform acts in England after 1832, largely as a result of the intensification of an enlightened humanitarianism. Earlier centuries had seen no similar concern with the material condition of the manual workers. A striking feature of the nineteenth century was its strong interest in social problems, and of this socialism was only one manifestation. Official investigations into working conditions, movements to protect child and female labor, strong interest by literary figures in reform—all revealed the humanitarianism of the time. But the socialists believed that the conditions which they deplored were not remediable under capitalism and that the creation of a new social system was necessary.

[2] See Chap. 2.

The Socialist Critique of Capitalism

Among the many schools of socialism there have been differences. Some have sought to extend government control over the economy and thereby enforce reforms; others have demanded that capitalism be overthrown by force. Neither have the objectives of socialists been uniform; some would strengthen the state and lessen individual economic if not political sovereignty within the framework of democratic institutions, and others would eventually abolish the state, or establish it as a dictatorship controlling all economic processes. But the socialist criticisms of capitalism have generally followed a familiar pattern, although from time to time one factor has been emphasized, then another.

First of all, goes the socialist critique, capitalism is unjust. There are rich and poor. As George Bernard Shaw, himself a socialist, said, at one end of the scale there are large dinners and no appetites and at the other end, large appetites and no dinners. The great inequality of incomes under capitalism does not accord, the socialists argue, with the equality of human beings. This does not mean that socialists want all individual incomes to be exactly equal. A few of them have desired this, but not many. Rather, most of them have insisted that incomes ought to be distributed according to the maxim, "from each according to his ability, to each according to his needs." This would mean that justice requires each individual to produce as he is able, but to draw income on the basis of his needs. Some socialists have concentrated their attacks on the extremes of inequality, insisting that incomes should reward effort expended for the common good and nothing more. Income inequality has, of course, been attacked by many groups other than socialists. The force of the socialist position is that it considers the inequality irremediable unless the system is discarded and a new and just system established.

Socialists have had much to say about the kinds of commodities produced under capitalism. They have pointed to the production of private yachts, of dog blankets, etc., at a time when many people lack adequate housing and children's clothing. This type of criticism is logically not separable from the objection to inequality of income, for in correspondence to unequal incomes there is the unequal distribution of consumers' demand. Since production under capitalism follows consumer demand (and costs), it is directed to the fuller satisfaction of the wants of those with the higher incomes.

A second major socialist criticism of capitalism centers about unemploy-

ment. A hundred years ago the concept of the business cycle was unknown. Depressions were recognized under the name of crises, and were regarded with less understanding if perhaps no less fear than they are now. The plight of the unemployed in periods of depression seemed particularly serious because as western Europe became urbanized the destitution resulting from unemployment assumed larger proportions. Unemployment in a system that was held to be wonderfully productive seemed senseless. Thus, capitalism has been indicted not only as inherently unjust but also unworkable, with unemployment offered as proof.

Third, in the twentieth century and especially since the early 1930's, socialists have laid a great deal of stress on the "unplanned" character of capitalism. This is partly an attack on the philosophy and practice of individualism, and may be in some instances a condemnation of political democracy. Competition as a method of organizing economic activity has always been viewed by socialists as the antithesis of proper organization of production and distribution. A standard socialist example of the folly of competition is the contrast between the mailman and the milkman. Look, says the socialist, at the efficiency of the government-operated Post Office Department: one mailman walks along the street leaving letters and parcels at their proper destinations. The milk distribution system epitomizes the duplications and wastes of competitive capitalism, because along the same street served by one letter carrier, three or four or even more milkmen travel, each one leaving bottles of milk at different doorways.

Furthermore, competition is criticized because it is believed to be a false and degrading principle: men should co-operate to carry out their common tasks rather than be opposed one to another for personal gain. Coupled with this there has been a general assault on profits and the profit motive. Like the word "capitalism" itself, the very word "profit" connotes evil to the socialist. Production, the socialists have cried for a hundred years and more, should be for use and not for profit. Profits and the profit motive are condemned on several grounds. The profit motive is a selfish one; individuals are forced to act in accordance with the principle of self-seeking gain, whereas the desire to serve others should properly be their aim. Under capitalism men produce whatever brings the greatest profits, not what their fellow men need. It must indeed be admitted that the socialists can cite a long list of examples of harmful and adulterated products that have been turned out under capitalism, such as a soothing sirup for babies which contained opium. To many sensitive persons in the nineteenth century, not all of them socialists, the commodities produced under capitalism were synonymous with cheapness and nastiness. Furthermore, the

profit motive is said to lead to the debasement of the people. Profit-seeking brewers, distillers, purveyors of bad literature and of degraded entertainment conspire to lower standards of taste. The drabness and monotony of modern life result under the profit system because the amelioration of living conditions is not directly a goal of employers. The efficiency of modern technology is thus believed to be perverted to bad ends.

The existence of profits is looked upon also by the socialist as the vehicle of exploitation. Profits are not earnings but plunder, they aver, taken from workers. The concept of the exploitation of labor involves more than the condemnation of profits. Logically, the objection to profits would come under four headings: (1) inequity exists because of high incomes, upon the assumption that individual capitalists' profits are always high; (2) profits are invariably looked upon by the socialists (and by some others) as an unearned and therefore undeserved income; (3) profits are seized from the laborer; (4) profits result from a selfish and hence false system of ethics. The exploitation of labor, as seen by early socialists, was associated with the conditions of work in the early factories. The factory system imposed a discipline on workers to which the twentieth century has become so accustomed that it is now scarcely noticed. The factory system a hundred years ago seemed harsh and inhuman: hours of work were very long, usually at least 12 hours a day 6 days a week; safety devices were scarcely heard of; corporal punishment and fines were imposed for violations of rules; and work was monotonous because the worker had to adjust himself to the unvarying rhythm of the steam engine. Working conditions in factories have of course improved tremendously in recent decades, but socialists are somewhat prone to speak of the past or concentrate attention on special instances of poor conditions which still exist.

Marxist Socialism

Of the many schools of socialism, by far the most important is associated with the name of Karl Marx (1818-1893). Marx's socialism is a fully developed economic thesis, and it has the greatest intellectual content. Marx has often been coupled with Darwin as a nineteenth-century figure who significantly influenced the course of modern thought. For it is the writings of Marx which has provided the doctrines of Europe's strongest socialist movements in the late nineteenth and early twentieth centuries, including that which culminated in the formation of the Soviet Union.

Karl Marx was born and educated in Germany. His father was a middle-class lawyer who hoped that his son would either follow in the law or

become a Prussian official. Instead, Marx became a revolutionary in the turbulent uprisings that spread over Europe in 1848, in Prussia, Belgium, Austria, and France. He fled to London in 1849 and remained there for the rest of his life in exile. After the failure of the 1848 revolutions Marx engaged in research, writing, and promoting the *First Internationale,* which he intended as an international league of workers.

The first of his writings to become widely known was *The Communist Manifesto,* published in 1848, which Marx wrote in collaboration with his friend and benefactor, Friedrich Engels, a successful textile manufacturer who had business interests in both Germany and England. *The Communist Manifesto* is an eloquent document and is still perhaps the best summary of Marxist principles. It is a ringing call to revolution which concludes by stating that the ends of the communists ". . . can be attained only by the forcible overthrow of all existing social conditions. Let the ruling classes tremble at a communist revolution. The proletarians have nothing to lose but their chains. They have a world to win. Workingmen of all countries, unite!"

"The Class Struggle Between Bourgeoisie and Proletariat"

The *Manifesto* begins by stating that "the history of all hitherto existing society is the history of class struggles," and this introduces a main Marxist idea, namely, *that capitalist and worker are bitterly and inalterably opposed to each other.* Capitalism—"modern bourgeois society"—is split up into "two great hostile camps, into two great classes directly facing each other—bourgeoisie and proletariat." Marx acknowledges the existence of other social groups, but insists that the only real division is that between capitalist and worker, the one the employing owner of the means of production, and the other the propertyless wage earner, and the real economic issue is the *class struggle.* The *Manifesto* continues with an account of the history of capitalism and of its "ruling class," the bourgeoisie. Capitalism arose out of feudalism, as a revolutionary change, the bourgeoisie having overthrown the power of the feudal landowners. Under capitalism it is the bourgeoisie who rule, for "the executive of the modern state is but a committee for managing the common affairs of the whole bourgeoisie." This is equivalent to saying that government is dominated exclusively by businessmen, and that labor has no voice at all. The bourgeoisie has accomplished much, continues the *Manifesto,* because it has revolutionized methods of production, but at the same time instead of the feudal "exploitation, veiled by religious and political illusions, it has substituted naked, shameless, direct

brutal exploitation." Furthermore, the system created by the bourgeoisie is destroying itself, because it suffers from successively "more extensive and more destructive crises," i.e., depressions, and because the exploited proletarians are becoming more numerous and more powerful. This happens because of the destruction of small business by capitalism, a process which forces more and more persons into the ranks of the proletariat. Capitalism rushes on to its own doom, until finally the whole system collapses. "The development of modern industry, therefore, cuts from under its feet the very foundation on which the bourgeoisie produces and appropriates products. What the bourgeoisie therefore produces, above all, are its own gravediggers. Its fall and the victory of the proletariat are equally inevitable." Then *The Communist Manifesto* proceeds to proclaim the communist party, in which Marx and Engels were members, as the only true representative of the proletariat, and to deny the validity of any other form of socialism or radicalism.

Marx called his socialist doctrine communism, and also scientific socialism. He did this to distinguish himself and his ideas from the other socialists of the period, whom he held in great contempt. Marx intended, by calling himself a scientific socialist, to show that for him socialism would come, not merely because anyone wanted it, but because of the inevitable processes of historical development. He sought to prove this in his other writings, especially in his great work, *Das Kapital* (Capital), Volume I of which was published in 1867; Volumes II and III were issued posthumously. *Das Kapital* is an economic treatise, an analysis of the functioning and dynamic development of capitalism. Its purpose is twofold, (1) to exhibit capitalism as inherently a system of exploitation, and (2) to prove that it is doomed to collapse of its own contradictions. Marx made a thorough study of the writings of the economists of his time and earlier, and his own theories are much influenced by the economists whom he was nevertheless wont to denounce as bourgeois apologists. Marx also investigated contemporary economic conditions, mainly through reading the reports of factory inspectors. Thus his thought was conditioned by both the ideas and the economic environment of the first half of the nineteenth century.

Surplus Value

His theory that capitalism is a system of exploitation can be briefly summarized. He begins with a labor theory of value; in modern terms this means a theory of the determination of prices, in which the key factor is the quantity of labor expended in the production of commodities. There

had been labor theories of value before, but Marx developed a theory for his particular purposes. The value of a commodity, Marx says, is created by human labor, and the quantity of value in a commodity is determined by the amount of labor contained in it. By value he does not mean price; price he terms *exchange value.* Value, for Marx, is the *normal* exchange ratio of a commodity, looked at apart from the vicissitudes of fluctuating market conditions. The value of a suit of clothes is, say, four times the value of a hat, because it takes four times as long to make the suit of clothes. To get around the problem of differences in the efficiency of labor, Marx defines the amount of value-creating labor as the amount that is "socially necessary," and this in turn is the amount of labor, working with machinery and performed with "the average degree of skill and intensity prevalent at the time."

With this qualification, then, Marx lays down his basic principle: value is created by human labor, and its amount is the amount of labor-time expended. For any given commodity, value should be measured from the beginning of the production process; the value of a suit of clothes, accordingly, is the labor-time required to produce the wool, the cloth, the other materials used, and including the wear and tear on the machinery employed as well as the final tailoring process. Marx takes it for granted that there is a demand for commodities; beyond this he made no special analysis of demand.

The process of capitalist production, according to Marx, is one in which the capitalist—the owner of means of production such as machinery, buildings, and raw material—buys a commodity called labor-power, produces values, and extracts from the whole process a surplus. The services of wage earners are labor-power, and labor-power is a commodity. Its value is determined in the same way as the value of any other commodity, i.e., by the amount of socially necessary labor-time contained in it. What Marx seeks to impress at this point is that wages are paid at subsistence levels. What the capitalist, or "moneybags" as Marx frequently termed him, buys in the free labor market is labor-power. From the exploitation of this labor-power the capitalist extracts his surplus. Labor-power is a unique commodity; it and it alone has the attribute of producing more value than is contained in it. The capitalist buys labor-power, paying, say, 6 shillings a day. But the labor-power produces, say, 12 shillings' worth of value in a day. The difference of 6 shillings a day is expropriated by the capitalist. This difference is *surplus value—the difference between what labor produces and what it is paid.* Surplus value is a fundamental concept in the Marxist theory.

Marx repeatedly asserts that the capitalist himself was not responsible for the exploitation of labor. It is the system which exploits; the individual capitalists only take advantage of the situation. The capitalist buys everything at its value—machinery, raw materials, and labor-power. He sells his product at its value. But he receives surplus value, because the system permits him to exploit labor, i.e., to extract surplus value. The capitalist himself produces nothing, he creates no values.

Marx's surplus value may be equated roughly with profits, when profits are taken to include all forms of property incomes such as interest and rent, as well as profits in the narrower sense. Profits come only from the exploitation of labor. Marx denied that profits arise in the process of exchange, for what one capitalist would gain in this way another capitalist would lose. For the system as a whole, profits are the sum of expropriated surplus value.

Marx's theory of capitalist exploitation of the wage earner is worked out in fine-spun and elaborate detail, but this much will perhaps suffice to continue a summary of his doctrine. Since surplus value is extracted from the laborer, it follows that the capitalist's interest is to exploit labor as much as possible. Hence Marx wrote of the efforts of the capitalists to lengthen the working day, because each additional hour brings in so much more surplus value. Furthermore, the capitalist seeks to speed up the pace of work, because in this manner also he can extract additional surplus value. It is to be repeated that Marx intended such charges as an indictment of the system: it is not ill-will or the chicanery of capitalists that leads them to oppress labor, but the system which by its very nature is exploitative.

Accumulation and Concentration of Capital and "Increasing Misery"

The second task which Marx set himself was the proof of the inevitability of the collapse of capitalism. From his theory of surplus value he proceeds to say that the additions to the total stock of capital goods come from the disposition of surplus value by the capitalists. The latters' own consumption comes parasitically from the surplus value they expropriate, but their consumption is less than the amount of surplus value which therefore becomes the source of the *accumulation of capital.* As a result of accumulation several things happen. One is the increasing *concentration of production* in fewer and larger capitalist enterprises, which is brought about because of technological advance, because of the competitive struggles among capitalists in which some are always being forced into the ranks of the proletariat, and because of the relentless pressure to invest surplus value. As accumulation and concentration take place, the "organic com-

position" of capital changes, i.e., wage expenditures become relatively smaller as compared with expenditures for capital goods. But wage expenditures are the only means of making a profit, of expropriating surplus value. Hence there is a tendency under capitalism for a "declining rate of profit" to set in. Marx defines the rate of profit as the ratio of surplus value to the total capital of the employer, whose total capital consists of two parts—labor-power employed, and all else. The proportion of the surplus value yielding outlays on labor-power steadily diminishes and with it the rate of profit. At the same time and for the same reasons, the position of the wage earner becomes worse. The capitalist can hope to stave off the decline in his profit rate only by oppressing labor still more. Thus wages are driven down, even below subsistence. The ranks of the unemployed, the "industrial reserve army," become larger. Hence, the "theory of increasing misery" of the proletariat. The depressions which rack the body of capitalism become more and more severe.

Thus, argues Marx, capitalism in its final stages develops contradictions which destroy it. There is the contradiction between the bourgeoisie and the proletariat, between the increasing productiveness of the system and the increasing misery of those who alone produce values, between the decline in the number of the capitalists and the growing numbers of the wage earners, between the method of extracting surplus value and the steadily mounting difficulty of doing so.

Dialectical Materialism and the Economic Interpretation of History

Finally the whole system breaks down completely, the communist party seizes power, the dictatorship of the proletariat comes into existence, and socialism emerges from the wreckage of capitalism. To understand why Marx believed this process is inevitable it is necessary to mention "dialectical materialism" and to comment on it briefly. "The dialectic," as it is known, is a theory of change which Marx adapted from the German philosopher Hegel, whom he had studied as a youth. Those who accept "the dialectic" believe that all change takes place through the juxtapositions and struggles between opposites or contradictions; that anything necessarily begets its own contradiction, or negation, and that the consequent struggle between contradictions culminates in a catastrophic transformation into a newer and higher form. To take the kind of example Marxist writers use: fire and water are opposites; they struggle until the temperature reaches 100° C, a catastrophic change occurs, and a new and higher thing called steam emerges. Marx subscribed also to a materialist philosophy: material

is the only reality, things of the mind being only a reflection of the material. For Marx, the essential material factor in all human life is the mode of production, for "forms of state [cannot] be understood by themselves, nor explained by the so-called general progress of the human mind. . . . The mode of production in material life determines the general character of the social, political, and spiritual processes of life." Hence, all historical change occurs through the dialectical movement of methods of production. This is at base an economic interpretation of history. Feudalism had developed its inner contradictions; a catastrophic change came about, and capitalism emerged. Capitalism has its contradictions, as the subject matter of three ponderous volumes of *Das Kapital* is aimed to prove, and capitalism must be destroyed by its inner conflicts—as anything must be—and a newer and higher form of society will emerge. Thus Marx was sure of the coming of socialism, and thus did he label himself a "scientific" socialist.

LENIN'S COMMUNISM

The picture that one gets from Marx's own extensive writings is that he expected capitalism *to end its own development,* and that this would occur in the *most advanced industrial country,* namely, England. However, at the beginning of the twentieth century capitalism was still strong and expanding, a fact which did not offer much comfort to Marx's disciples. One of these was the Russian revolutionary Lenin, who in his writing developed some of Marx's ideas with the purpose of showing the possibility of the outbreak of "the" revolution in Russia, rather than in the advanced industrial countries.

Lenin headed the Marxist party known as the "Bolsheviks" which in 1917 overthrew the Provisional government set up following the fall of the Czar in the Russian Revolution. Lenin was leader of the new Soviet government until his death in 1924, following which Stalin forced Lenin's first associate and co-theorist, Trotsky, from power and became the ruler of Russia. Russian communism's greatest intellectual thinkers were Lenin and Trotsky; however, it is Lenin's Marxian thinking which serves as the official philosophy of the Soviet Union. Trotsky was discredited as a result of Stalin's political ascent and "Trotskyism" is now bitterly opposed by the Soviet communists today.

Finance Capitalism and Imperialism—the Later Stage

Lenin developed a theory of "finance capitalism" and of "imperialism" in order to take into account the political and economic changes since Marx's time, as well as to show the possibility of revolution in Russia. According to Lenin, the evolution of capitalism had proceeded to the stage where, in the early twentieth century, economic and therefore political domination was exercised by investment bankers. Marx's capitalist had been an employer-owner, an industrialist, a competitively operating, somewhat small businessman. Lenin's capitalist was a monopolistic financier controlling many factories. Communists still use Lenin's expressions "finance capitalism" and "monopoly capitalism" almost interchangeably to describe the twentieth-century economy. Lenin also gave his attention to the imperialist activities of the major European nations in Africa and Asia in the late nineteenth and early twentieth centuries. These were the export of capital to backward areas, the search for raw materials, and the "oppression and exploitation of colonial peoples." The finance capitalists used their governments' armies and navies in ruthless attempts to increase surplus value by colonial exploitation. The struggles within industrial capitalism had become transformed into a world-wide struggle between finance capitalists on one side, and the European proletariat and oppressed colonial peoples on the other side. Most important of all, there was a struggle among the finance capitalists themselves on a national basis. Hence, argued Lenin, war was inevitable; with the outbreak of war, revolution would occur, and it would take place at the weakest point in the "chain of world imperialism," namely, in Russia.

Dictatorship of the Proletariat and the Removal of Government

Marx had written very little on the subject of how a socialist system would operate. Lenin followed up some brief suggestions of Marx, amplified them, and before the Russian revolution occurred gave the following picture of what was to come with the collapse of capitalism. The communist party would seize power, and establish the dictatorship of the proletariat, using *soviets* (councils) as administrative vehicles. The soviets were to be composed of "workers and peasants" and would serve under the leadership of the communist party. The first task would be to establish socialism: the ownership and operation of the means of production by the victorious dictatorship of the proletariat. Under socialism, incomes would be paid accord-

ing to effort. Government, which is necessarily and always an instrument of oppression, would be employed to liquidate the remnants of the former ruling class. This would be the first stage—the stage of socialism, or the lower phase of communism.

The second or higher phase of communism, or pure communism, would come about presumably after the world-wide success of the communist party, and after the disposal of all possible dissidents. In this stage "the state would wither away," because there being no longer anyone to oppress, there would be no need for the state and its apparatus of oppression. Pure communism would also be the "classless society." Production would become so bountiful that there could no longer be any quarrels over the ownership of economic goods and hence no classes.

CRITICISM OF MARXIAN TEACHING

What can be said in criticism of the ideas of Marx and Lenin? It is well to remember that these ideas have been and are being taught as eternal verities to millions upon millions of people, to whom all opposing ideas are worse than wrong.

To begin with, criticism can be leveled at Marx's theory of value. It is a type of theory long since discarded by economists, because at best it is only an effort to explain prices by reference solely to costs of production. From the modern standpoint, it is an old-fashioned theory, crude, and with a misplaced emphasis. It is not completely wrong, but it is no longer useful. Second, criticism of Marx's doctrine of surplus value is possible. Marx made this hang on the uniqueness of labor-power, the only commodity capable of producing more than it cost. Logically, one could pick out any commodity and make the same declaration. Wheat, or iron, or coal, or machinery would be examples; and the owners of these could be shown in the same way to be "exploited." Furthermore, Marx's class-struggle theory is too simple, in its division of all persons into two opposing groups. Where does a salaried corporation executive fit? Is he bourgeois? He cannot be, for most likely he owns no means of production. Is he proletarian? He is a wage earner and is therefore "exploited"; but after all, is the salaried executive ground down, oppressed and eking out a bare subsistence income? And what of the independent professional people, the government employees, the farmer who owns his own land and equipment? Where do they fit in the Marxist system?

The evolution of capitalism since the middle of the nineteenth century has borne out Marx's predictions only in small part. One of Marx's genuine

contributions to economic knowledge was his theory of concentration. However, Marx claimed that concentration is a general tendency, and he saw large units arising in all parts of the economy. But small units continue to the present day in such activities as agriculture, clothing manufacture, and printing. Nearly a hundred years have gone by since Marx promulgated his laws of the declining rate of profit and the increasing *misery* of the proletariat. Profits have risen and fallen, and it would be hard to find the alleged long-run tendency to decline. The position of wage earners under capitalism, it is scarcely necessary to add, has vastly improved since the middle of the nineteenth century.

In fairness to Marx it must be pointed out that his description of capitalism fitted the conditions of his time in many respects. Since then many changes have taken place in the economic system called capitalism. Unions have assumed a significant place in the economy, labor and social security legislation have done much to protect the wage earner, and government control has extended itself over many areas. Marx's own conception of capitalism was in many ways a one-sided version of the system of his time; if examined literally today, it can only be a ridiculous caricature of the modern economy.

The geographical area where capitalism now prevails is indeed smaller than it was, say, in the year 1900. This has been due to the slow growth of government controls and above all to the profound *political* changes brought about by the two world wars. Capitalism can exist only in a society characterized by political stability, by a government of law, operating within a democratic framework.

OTHER SCHOOLS OF SOCIALISM

Utopian Socialists

The French and British socialists of the late eighteenth century and in the period before 1848 are customarily regarded as utopian socialists, because they devoted most of their efforts to describing and promoting societies to be perfected on unselfish communal principles. Some of them led colonizing expeditions to North America to establish model communities in the pure air of the New World away from the vices and corruptions of Europe. Among these were the Utica Community and the New Harmony in Indiana, the latter led by Robert Owen. It was hoped that men could work together prosperously and happily in equalitarian communities according to simple economic plans which eliminated individual gain from production. The utopian socialists sought to achieve their ends by appeals to reason and by

persuasion and had in fact a profound horror of violence; many were active religious leaders and drew on Christian teaching to support their principles.

Fabian Socialists

In the 1880's there formed in England a group of persons who called themselves Fabian socialists. They included many distinguished writers, such as George Bernard Shaw, H. G. Wells, and Sidney and Beatrice Webb. They chose the name of "The Fabian Society" because Fabius, the Roman general who successfully opposed Hannibal, had employed delaying tactics until he wore out his adversary, waiting until the right moment to strike the decisive blow. The Fabian socialists wrote series of essays and books of considerable if not dramatic influence, especially in forming the ideas of the British Labour party, many of whose intellectual elite were either members of the Fabian Society or closely associated with it.

The Fabians believed in the inevitability of the coming of socialism, although their reasons were different from those of Marx. The Fabians thought that the slow growth of governmental control and of government ownership, by a kind of Darwinian evolutionary process, would finally eventuate in socialism. Step by step, the government extends its influence, until at last it must become all-embracing. In the period from 1890-1910, when the modern utility industries—electricity, street railway, water, and gas facilities—became fully developed, many cities of Europe placed them under municipal ownership. The Fabians greeted this development with enthusiasm, and indeed were for a time known as the "gas-and-water socialists." The Fabians stressed the injustices of capitalism, particularly the inequality in the distribution of income. Among the concepts they borrowed from the writings of economists was their belief in the necessity for doing away with private receipt of the rent from land. For this they owed much to the work of an American: Henry George's *Progress and Poverty* (1879). George, using the early nineteenth-century theories of rent, wages, and profits, argued that most economic ills are due to the private ownership of land. George held that as labor increases its productivity the landowners take the gains in rent. Rent is derived from the value which the community as a whole creates. The gains belong, then, to the community, not to landowners who at present enjoy the unearned income. George advocated the abolition of the private ownership of land, and proposed that this be done by a confiscatory tax on rent, a tax which would render unnecessary all other taxes, i.e., *a single tax.*

The Fabians, as good Englishmen, have been firm and sincere believers

in democracy and peaceful political change. In this respect, too, they differ sharply from Marx and his followers. Socialism in the Fabian view would come about through act of Parliament. (It is too early to judge the success, or lack of it, of the measures adopted by the British Labour government voted into office in 1945.)

Revisionists

About the same time that the Fabians proposed their socialism, some of the followers of Marx were modifying his ideas and developing less violent programs of political action. These were the "Revisionists," or Social Democrats. Marx was a revolutionary, but the revisionists saw the possibility of a peaceful, evolutionary, democratic march to socialism. The Social Democratic party gained seats in the German Reichstag; they participated in legislative procedure and sought thereby to outvote the bourgeois power that Marx would have attacked with violence; nevertheless they considered themselves Marxian socialists. With the Bolshevist revolution in Russia, the Marxists in other countries split into two main groups, the communists proper and the social democrats. The former have followed the line of Marx and Lenin and have largely been controlled by Russia's policy of international communism; the latter, although acknowledging inspiration from Marx, have worked for socialist ascendancy within their own countries and by means of the democratic principle.

Syndicalists

Another type of socialism which came into existence in the beginning of the twentieth century was syndicalism. This was a fusion of some of Marx's ideas and slogans and of trade unionism. The word "syndicalism" comes from the French word for union, *syndicat*. The syndicalist movement flourished mainly in the Latin countries—France, Italy, and Spain—although a version of it was widely discussed for a while in England under the name of "guild socialism." In America, syndicalism seems to have emerged spontaneously. In this country it was represented by an organization known as the "I.W.W.," the Industrial Workers of the World (known to their opponents as the "Wobblies"), which existed in the West chiefly among loggers, miners, and migratory farm workers. The leading idea of syndicalism is that the union is at the same time both the true instrument for the destruction of capitalism and the social unit or cell of future syndicalist society. The syndicalists deplored all forms of government as being a class instrument for

oppression; in a syndicalist society there would be no organized government but only workers' organizations loosely co-operating with one another. The syndicalists believed in violent attack on government, chiefly by means of strikes. In France it was these socialists who spread the word "sabotage," which is derived from the sometime use of wooden shoes (*sabot*) to jam machinery during strike uprisings. Even today many of the states in this country have laws against "criminal syndicalism" which were passed in the period before 1920, when the I.W.W. carried on a campaign marked by some violence. The syndicalists harassed employers for the sake of doing so and for the practice in revolutionary tactics. A strike lost was a strike won for the revolution; they preferred not to attain higher wages and better working conditions, for such achievements might bring even temporary contentment to the workers.

Syndicalism mostly disappeared in the period just after World War I. Communist parties in the various countries absorbed many syndicalist party members and took over the torch of the tradition of Marxist violence.

Socialist Conflicts

The main tendencies and conflicts of socioeconomic programs at the present time can be summed up by reference to two factors: the attitude toward private enterprise, or capitalism, and the attitude toward democratic government. The communists are of course mortally opposed to capitalism, and their concept of democracy, a word which they use frequently in describing Soviet practices, is not the traditional one of Western political history, for it denies most individual civil rights. Social democrats believe in government ownership of industry within the framework of democratic government. Support of private enterprise is necessarily accompanied by adherence to the notion of democratic government. In each one of these programs many qualifications must be made. Among communists, for example, some have been followers of Trotsky rather than of Stalin; the nominal difference between them was concerned with the rate of speed at which world revolution should be fomented, the immediate dispute centering upon control of power in Russia. Among social democrats too there has been much quarreling over principle and considerable deviation in the degree to which some have been willing to compromise with capitalist methods. Capitalism itself has, moreover, been given to many divergent points of view, as was noted in Chapter 7.

FASCISM

The fascist governments of Germany and Italy were deposed in World War II but the fascist ideology is not yet dead. It is difficult to define fascism because its forms and sources have varied from time to time and from country to country. Germany and Italy provide the clearest examples.

On the level of propaganda, German and Italian fascism was avowedly anticapitalistic. This was due in part to the necessity of appealing to masses of people who had long been conditioned by socialist propaganda. Indeed, the nazi party was officially the "National Socialist German Workers' Party." The economic systems of both Italy and Germany were disrupted by World War I; revolt was widespread, and nationalistic feelings were fanned. What the fascist parties promised was a new economic and a new political system.

The ideas concerning the proposed fascist system were worked out much less fully in Germany than in Italy where they were centered around the concept of "the corporate state." This Mussolini once described as a higher synthesis of individualism and communism. The corporate state existed in Italy in actuality only as a paper organization. But it was intended to organize all producers, whether businessmen or wage earners, in hierarchical groups called corporations which in turn were to be integrated into a national system under the dominance and leadership of the Fascist party.

In practice, fascism in Italy and Germany was a system in which the fascist and the nazi parties exercised total control over the economy for political and avowedly military ends. Fascism is a totalitarian system: power is invested in a few and the political rights of the many are denied. While private property is not abolished, property rights were not respected outside the political conformity of property owners. Civil privileges depend on conformity to the philosophy of the state, which varied among fascist nations. Fascism has in general been supported by their armies and big business, although again it is difficult to determine at what time and in what circumstances such support is forthcoming in a system like fascism which depends so greatly on the prejudices of a dictator clique. Fascism differs from the socialism practiced in Russia mainly in its retention of private ownership of economic resources. It resembles the system in Russia in its totalitarianism and police rule, in the use of the party-state form of government, and in the attitudes and practices toward dissidents. Although there is a similarity between fascism and democratic capitalism in the operation of private property functions, such as private securities markets and banking, the supremacy of the state in the former is seen in the fascist

theory that the individual can hold private property only as a kind of trustee for the nation, and in the fascist practice of large-scale confiscations of property when it suited the purposes of the dictatorship.

Fascism is a form of collectivism. Collectivism is a social system in which men and resources are mobilized in the service of a stated purpose of the government to which everything is subordinated. Socialism is therefore also a form of collectivism. Marxists have insisted that fascism is a decadent version of capitalism, and fiercely resent their system's being bracketed along with fascism. On the other hand, many social democrats have found the lack of democracy under both fascism and communism sufficient reason to lump the two together.

The lesson to be learned from the experience of Europe with fascism is that collectivism is possible without democracy and without the formal abandonment of private property in economic resources. After all, the fascist and nazi organizations did achieve what socialists had long demanded, namely, the completest degree of governmental control.

The Issues Raised by Collectivism

Socialism as an idea has been an aspiration for a better world. There are two serious questions which can be raised in an examination of socialism. One is whether the particular ends of the socialists are good, and the other is whether the socialists have proposed the correct means for attaining their ends. It is well to recognize some of the obviously fallacious theory that has been advanced by them. An important example of this is their promise of abundance. If it means anything at all, abundance is a relation between goods and wants such that all wants are satisfied. This is not absolutely possible even for a community of ascetics. If abundance simply means more than is produced per capita under the present economic system, then it becomes a question of judging probabilities. During the depression of the 1930's, the advocates of over-all economic planning (which is equivalent to socialism) made many extravagant claims as to how much total output would be increased if only co-ordinated government planning were effected. The recent war showed how much the present economy can produce. Real output was a little more than twice as high as during the depression. The withdrawal of men into the armed services was approximately offset by the entrance into the labor force of persons not normally a part of it and by the increase of hours of work. Moreover, the war saw the intensification of effort and the stimulation of incentives which would scarcely be possible continuously in peacetime. But war output was not three or four or ten

times as high as peacetime output; yet socialists had promised increases in output of at least this much if socialist planning were adopted.

Socialism, whatever its merits, cannot be looked upon as an organizing principle which would increase total output. Probably, the convinced socialist would not concern himself too much with this matter anyway, for if offered a choice between capitalism and a socialism with a definitely lower output, he would undoubtedly choose the latter. The basic purpose of socialism is the organization of life on some collective principle: the dictatorship of the proletariat of the communists, the syndical society of the syndicalists, the state ownership of key resources of the social democrats. The contrast between all of these and the present economic system is a contrast between collectivism and a system which is still predominantly individualist despite the large measure of government control and the existence of highly organized groups. The end proposed by the socialists is the collectivization of life. That this is desirable is surely disputable. The means proposed is government ownership and operation of all or nearly all economic resources. Experience in the twentieth century ought to cast doubt on the beneficence of activities carried out by governments, even by the most democratic.

Socialists are likely to offer themselves as the true representatives of economic democracy. Some of them in fact have defined socialism as economic democracy. Now what does economic democracy mean? Unless it is to be merely a pleasant-sounding phrase, economic democracy would have to mean that every production decision would have to be made by vote of the workers concerned. This situation would mean such a chaos that thoughtful socialists have shied away from it. Economic democracy can therefore mean only that an elected government operates the economy by administrative regulation; and this is no different from saying that political democracy ought to seize the major economic resources and administer them.

CAPITALISM AND DEMOCRACY

A word or two on capitalism and democracy is in place here. It is no historical accident that capitalism and free civil government are found together, both in time and in place. Both are manifestations of the same social principles: the supremacy of the individual, government of law, and the right of free association.

A much and bitterly disputed question is the compatibility of socialism and democracy. The communists are under no illusions on this point. The writings of Lenin are quite clear in stating that socialism must be a *dictator-*

ship of the proletariat, it must carry out oppression of the remnants of the former ruling classes. What is ordinarily known as democracy Lenin held in contempt; he called it a fraud and a sham perpetrated on the toiling masses. Marx and Lenin both regarded all government as simply an instrument of oppression and believed that only when government is abolished are men truly free and happy. But other socialists have believed in parliamentary processes of orderly government; and a central question therefore is whether this is possible under socialism.

That socialism and democracy can exist together can be denied for several reasons. Political democracy has existed successfully in a context where the powers of government are limited and where both political parties agree on main principles. A socialist democracy would invest total control over the economy in a legislature. This would in turn mean the impossibility of effective planning, because an over-all economic plan must represent one idea, one integrated set of purposes. A legislature cannot draw up such a plan, unless most of the members subordinate themselves to a strong leadership. If strong leaders should emerge to force a plan on all of society, they would have an amount of power which no politician in a capitalist democracy has ever had, even in wartime. The temptation to wield and to keep this power would likely be so overwhelming as to destroy political democracy, even if its outward forms were kept alive. "Power corrupts; absolute power corrupts absolutely."

CAPITALISM AND SECURITY

Socialism also raises the age-old question of freedom versus security. A *laissez-faire* society—"anarchy plus the constable," Carlyle termed it—represents the maximum degree of freedom in the modern age; America in the early part of the nineteenth century closely approximated this state of affairs. This high degree of freedom is accompanied by a minimum of—organized—security. Security of a sort has never been completely lacking under capitalism; the freedom to change jobs is one of the major aspects of security. Poor relief is an ancient institution, and while mostly it has been inadequate it has been there. The opposite end of the spectrum of socialist systems is probably occupied by the dictatorship of the proletariat, with a minimum of freedom—for the individual—and a maximum of organized security. In between *laissez faire* and the dictatorship of the proletariat lie many possibilities of combining security and freedom. The socialists always promise both and more of each than they say exists under capitalism. No doubt security can be provided in systems other than capitalism, but the

real issue is whether socialism could also afford freedom to the individual. A socialist system would necessarily control resources, and would control them under central direction and according to a plan. This would entail a control over production which would affect the individual as a worker and would probably give him less freedom of choice of occupation and less freedom to change jobs than he has under capitalism. The socialist control over the production and distribution of consumer goods means ultimately that the individual does what government authorities want him to do and consumes what the authorities decide is good for him.

CHAPTER 42

The Economic Systems

Undoubtedly, one reason for the spread of socialist doctrines in the last several decades is that socialism as a set of *ideas* has been constantly compared with the *facts* of an established capitalism. It is likely that a doctrinaire promise of a better world will be more appealing than the reality of any actually functioning economic system. Were the tables turned, and if in a socialist society it were possible to propagandize freely for capitalism, then capitalism would be presented in a most favorable light in sharp contrast to the drab realities of the socialist world. Soon, no doubt, a subversive capitalist movement would arise, still on the supposition that the socialist state would permit such activities.

ECONOMIC FACTORS

Socialism in practice is bound to be disappointing to the socialists. And this not simply because expectation is better than fulfillment, but because of certain basic and unavoidable economic factors. These can be examined briefly.

Scarcity

Socialists have long described capitalism as a system of scarcity, one which breeds scarcities. Friedrich Engels, Marx's collaborator, said that the coming of socialism would mean for mankind a leap from the realm of necessity into the realm of freedom. But economic scarcity exists in all economic systems, including the socialist, simply because it is the relation between wants and resources. The whole tendency of the modern age has been for wants to expand faster than the means of satisfying them. For the very nature of expanding and refining productive facilities—transportation, manufacturing processes, mining, agriculture, etc.—is that old wants are stimulated and new wants are created. Human wants are not, however, uncontrollable; it must be recognized that it is possible to confine a people's wants within certain limits by force or by appeals to their loyalty to a cause and nationalistic feelings. In modern times this restriction of wants has been demonstrated by the British Labour government's "austerity" program and by the Russian Five-Year Plans, in which each government seeks to withhold production of consumer goods to maintain high production of capital facilities. It must therefore be admitted that modern propaganda, particularly when, as in Russia, it flows from only one source, can accomplish the task of basing a national policy on individuals' poverty. Yet this accomplishment, of whatever virtue it may be to state plans, reveals that the fact of scarcity is the fact of economy and that it will obtain in a socialist economy as inevitably as in a capitalist.

Economizing

The first requisite of the socialist state, then, is to economize: to manage scarce resources in order to attain its ends. This will mean work. The visions of the old-fashioned socialists of a four-hour day must remain very distant. Economizing in the socialist state will also mean a decision between present and future consumption, i.e., a decision as to what proportion of total national income is to be saved. Some such decision will have to be made unless the socialist state is to use up whatever capital equipment it would start with, and so eventually impoverish itself. Part of current effort therefore will have to be devoted to the maintenance and increase of capital equipment. Thus instead of providing more consumer goods, the contemporary socialist systems must ask—or demand—sacrifices from the population.

Furthermore, to continue with the economic needs of a socialist system,

it is obvious that a socialist economic system must provide a mechanism for directing production. It is not enough to say, although most socialists have in some way implied it, that "common ownership of the means of production" will of itself simply produce the things which "the people need." Who knows what the people want? The socialist is impatient of the question, taking it for granted that anyone knows that the primary wants are food, clothing, and shelter. But still, it can be asked, what kinds of food and in what proportions? Whose dietary standards are to be followed? Who is to get the better cuts of meat, and who the plate beef? Who gets which qualities of what particular foodstuffs? And as for housing, which kinds of houses are to be built where? Who gets the new dwellings and who must live in the old ones? Likewise for clothing. These problems could be solved very directly of course by producing a standard food ration for an entire population which would be housed in uniform barracks and dressed in uniform clothing. But this would hardly be the promised land of socialism; it is difficult to imagine anyone longing for such a nightmare.

Economic Decisions

Socialism must therefore devise a system in which decisions are made on the following problems:

1. The kinds of *particular* goods and services to be produced.
2. The proportions in which the goods are to be produced. If resources are available for more houses or more office buildings, but not both, for example, it will have to be decided which are to be constructed, and how labor and materials are to be procured, and how labor and materials are to be divided between them.
3. The various methods of production. It is again not enough to say, as some socialists have, that the latest and most efficient methods will be employed. Technical efficiency is an engineering criterion and cannot be considered apart from basic economic decisions, such as the relative scarcities of the different grades of labor and of different natural resources available, the productivity of old but still usable capital goods, and the scarcity of new capital.
4. The distribution of income to be effected and the methods by which to distribute income. (This last point shall be discussed later.)

Moreover the production problems of a socialist will involve not only making decisions of the kind described but also of devising some sort of a system or mechanism by which they will be put into effect. The resources

and the wants of an isolated rural community or of a primitive tribe can probably be encompassed by a single mind. Such simple communities almost always live and work within a rigid framework of custom, a fact which greatly simplifies the process of making economic (as well as social and political) decisions. But the modern economy is so vastly intricate, so finely adjusted, and so little bound by rigid custom that no single mind or even a group intelligence can absorb all the pertinent facts and co-ordinate them into a plan for production. In allocating goods among the millions of consumers, for example, assume that the purpose of the allocation is to meet consumers' wants as the consumers themselves interpret them. This can be tried by putting prices on the goods, providing the consumers with different means to obtain money, and allowing them to buy what they want and what they can. This is substantially what the Russians do, and modern socialists who know something of economics advocate such a practice for the economies they propose. The alternative is to give everyone a ration which will be allocated to persons on the decision of some administrative authority.

Money and prices, therefore, would have to be part of the mechanism. Money indeed would be essential throughout, for it is the only way in which comparisons (exchange) can be made. Some socialists have advocated management "in kind," the way, they said, "the army does it." What they meant was that thinking at the top decision level should be in tons, gallons, board feet, etc., in the physical units in which commodities come. Thinking in terms of money smacked to them too much of capitalism. But if decisions were to be made only in terms of physical quantities of commodities, how, for example, would the importance of a million board feet of a certain kind of lumber be compared with the importance of fifty thousand tons of potatoes? The difficulties of primitive barter would be nothing compared with trying to make decisions in terms of physical commodities. Money is necessary as a means of comparison; it expresses value, i.e., the importance of one commodity in relation to others.

In any economy, once the values of resources and the available technical methods of production are known, it is possible rationally to allocate resources so as to maximize the satisfaction of wants, whatever they are. The problem of a socialist society therefore is to determine the values of the resources, material and human. This is at best a difficult problem, one on which there has been much controversy among economic theorists, some of whom have denied that a socialist society could devise a mechanism to determine values at all, except purely arbitrarily. If values were determined arbitrarily, the socialist society might produce less than it conceivably could, and no one would ever know whether or not production was directed

for the right things in the right proportions, i.e., for the maximization of want satisfaction. Other economic theorists have said that a socialist economy could actually solve the problem of determining valid indexes of importance for its resources, by creating within its framework a market mechanism identical with that of capitalism.

The problem of production decisions in a socialist state becomes much simpler if the wants to be satisfied are social objectives of overriding urgency, rather than wants which themselves have to be determined from the vast sum of individual preferences. If a socialist economy is engaged in fighting or actively preparing for a war, its purposes are of necessity military and are accepted by those in charge of making the production decisions. If too the socialist economy is directed toward the execution of vast developmental projects, to which everything that can be is subordinated, the economic authorities simply carry out the ends of the higher political authorities. In both cases, political decisions determine the importance, the effective values, of the various end products of the economy. But where the wants themselves have also to be determined, as is the case if human welfare were to be the goal of socialism, the task of directing production becomes much more difficult.

The distribution of income also brings a socialist economy face to face with stark economic necessities. Socialists have always hoped for a society in which incomes would be distributed justly. Standards of justice proposed by the socialists have varied, but there has been at least one thing which socialists have in common: incomes are to be much less unequal than under capitalism. To the stock objection—who would perform the disagreeable tasks under socialism?—the reply has usually been that those engaged in the less pleasant jobs would be rewarded by shorter hours of work than the average, or by slightly higher pay, or by some combination of both.

The substantial equality of incomes which the socialists have advocated is incompatible with the requirements of labor allocation and with the existence of an effective incentives system. Consider first the problem of allocating labor—among the various occupations, among industries, and among geographical areas. Suppose that in a socialist country it is decided to construct a new aircraft manufacturing plant in an area where at present the labor supply is inadequate both in number and in composition. The workers needed can be trained or shifted from other industries; in any case they will have to move from their present homes to the area where the new plant is situated. How can the right number of the right kinds of workers be moved to the new plant? Compulsion is one method. Persuasion is another but probably ineffective one. The simplest way would be to offer

higher wages in the new plant. If compulsion is employed, equality of incomes could be maintained, but a society which would employ force is not one which would attach much weight to the principle of the equality of men. If the workers are free to work when they please, they will not go to the right places if incomes are more or less equal. In brief, a socialist society would be forced either to compel its workers to allocate themselves in accordance with the production plan, or to manipulate wages to bring about the same result. The wage manipulations necessary to ensure proper labor allocation would not be slight. They would be large enough to make a mockery of efforts to maintain equality of incomes.

These considerations are one reason that a socialist society could not tolerate wage bargaining by workers' organizations. There could be unions in a socialist state, but the unions would have to limit their activities to the handling of minor grievances.

For workers in given jobs, incentives are necessary. This is a complex problem because the incentives to human effort are varied. The joy in work itself, the pride of accomplishment, the desire of approbation are indeed often powerful factors; and the socialists have hoped that these forces would provide the stimuli to higher output. But differentiated money incomes are an incentive too, the most powerful single incentive for most people most of the time. Production is not so large when the workers receive relatively equal incomes as when incomes are differentiated, as the experience of the Soviet Union proves.

THE EVOLUTION OF SOCIALISM IN RUSSIA

The Soviet Union is a socialist country. It does not practice communism, in the sense that the Russians themselves interpret it from the writ of Marx and Lenin, which defines socialism as a beginning stage and communism as a final stage of development. But since the government of Russia owns and operates all resources, the present position of Russia fits just as well the non-Marxist definition of socialism.

Socialism as practiced in Russia is in some respects so different from many important socialist ideals that there has been a degree of backsliding among socialists in other countries. If this is socialism, they have said in effect, let us have none of it. Most socialists, however, cling to a belief in "the social ownership of the means of production"; so long as this is in effect, all is well to them, whatever appearances to the contrary may be.

Socialism in Russia came into effect in a technically backward country with a tradition of oppressive government, a largely illiterate population, a

very small amount of capitalist production, and a society torn by war and civil war. Circumstances could scarcely have been more unfavorable. For this reason, much that has happened in Russia in the last quarter of a century is irrelevant, in the sense that it affords no useful lesson to the advanced Western nations. Comments on the Soviet Union will be confined to those matters which pertain to its specifically socialist character, although it is realized that it is often difficult to separate that which is socialist from that which is Russian.

Marx wrote many volumes on the economy of capitalism, but almost nothing on how a socialist economy would operate. Lenin also largely confined himself to economic and social dogma. Consequently, when the Bolsheviks seized power in 1917 they had no actual measures planned on how to establish the economics of their new dictatorship of the proletariat. They toyed briefly with the notion of using Marx's labor theory of value as a guide to production but quickly gave it up. But what they lacked in economic ideas the Bolsheviks made up in political skill and their ability to retain power. Over the course of the years the present system evolved. Much that the Russian leaders have learned through the costly method of experience they could have found out by systematic study of economics, but they believe that all economic literature, save that of Marx and of his approved followers, is so much bourgeois, reactionary claptrap.

The first important step in the development of socialism in Russia was "War Communism," in the period 1918-21. All resources except agricultural land were appropriated for governmental operation, but no special system or mechanism of production was devised. Production was determined administratively, and in terms of physical units rather than in money units. This was the method of "natural socialism." However, War Communism did not prove successful, partly because the economy was disrupted by civil war but mainly because it allowed production to fall to a disastrously low level. Finally the system under War Communism was recognized as faulty and abandoned.

The next development was the New Economic Policy, begun by Lenin in 1921. The government continued to own and operate the railroads, the banks, the foreign trade organization, and the key industries, but beyond these permitted private enterprise and private trading. Agricultural land continued in private ownership, largely in the hands of the peasants, many of whom had benefited from the division of large estates. Businessmen came out of hiding, trading in commodities was resumed, farm products appeared in much larger quantities in the cities, and many kinds of small business were re-established or initiated. Under this mixture of socialist

and private enterprise in the New Economic Policy the population prospered in comparison to their status under War Communism. Neither under War Communism nor under the New Economic Policy was an over-all plan in effect. The idea of planning (i.e., a mechanism of control) emerged slowly, and appears to have arisen in connection with the making of statistical estimates of the production of peasant farmers, of the small businessmen who were allowed to exist for a time, and of the state enterprises. The statistical estimates came in the form of "control-figures," which were intended at first not to direct production but to afford suggestions to the heads of the state enterprises.

Difficulties arose in the middle 1920's over the procurement of foodstuffs for consumption in cities and of grain for export. The peasants would not sell their products unless they obtained prices which they deemed satisfactory in relation to the prices they had to pay for industrial goods. The prices they paid were set in two ways: the administered prices of the state enterprises, and the market prices of the private segment of the economy. The latter were more favorable. As a consequence, the deliveries of farm products did not meet the demands of the authorities, and a considerable impatience with the peasants was manifested in high circles of the communist party. After Stalin's supremacy in the party was firmly established it was decided to set up a functioning plan, and thus the fate of private enterprise was sealed. Private trading in grain was forbidden, private enterprises in the cities were abolished, and the small businessmen were liquidated. Replacing the New Economic Policy, the first Five-Year Plan was put into effect in 1929, to operate as from 1928.

During the period of the first Five-Year Plan the problem of the individualistically inclined peasants was solved. Private ownership of land was abolished; every farmer who was not already working on a state farm was forced into collective farms, which are large-scale co-operative enterprises. Marx had defined exploitation as essentially the employment of one person by another. With unimportant exceptions, no one in Russia has employed anyone else since the early 1930's. This aspect of the Russian system is not so much an economic decision as it is slavish adherence to dogma.

The purposes of the first Five-Year Plan were to increase production (especially in the capital goods industries), to co-ordinate all production, to develop new resources, to solve the problem of getting foodstuffs from the peasants, and "to ensure the victory of the socialist economic system in its historic contest with capitalism." The rate of investment planned and actually carried through was very high, much higher than the proportion of capital goods production to total output than has ever obtained in a

capitalist country in peacetime. The original plan was to accomplish this without reducing the volume of consumption goods output; in fact it was planned to increase the latter. This seems not to have been accomplished, however; the standard of living actually fell somewhat, according to what information is available.

Perhaps the best way to visualize the nature of the Russian five-year planning system is as a permanent war economy. There is the same singleness of purpose, the relentless mobilization of all men and all resources for one unchallenged cause, the same whipping up of the population's enthusiasm.

Planning on the scale practiced in Russia since 1928 affects even the smallest details in the economy. Manufacturing, transportation, mining, domestic trade are co-ordinated by a central plan on which the principal production decisions for the entire economy are made. Each major sector of the economy has its central trust, directing the activities of its regional subsidiaries. For each plant, mill, mine, and farm there is a production quota set by the plan of its trust, which is in turn governed by the master plan. Each unit in this integrated production scheme is under constant compulsion to meet and possibly exceed the established quota; thus, an increase in one unit then calls for increases in related units, otherwise production becomes disjointed—engine blocks may, for example, exceed the supply of component parts for the engines. The Five-Year Plans thus are modified *in detail* constantly; production decisions are revised in an effort to produce the vast number of different things in the right amounts at the right time.

A planning system of this type presupposes a given scale of preferences, i.e., a given hierarchy of wants, otherwise centralized production decisions would not be possible. The planners know what they want. Their main problem is to apportion scarce means to satisfy the scale of preferences they or their political leaders impose on the country. It is important to note that the given scale of preferences must be centrally established, and the disposition of means must also be centrally regulated. In contrast, the scale of preferences in capitalism is not centrally determined; the preferences of each person and family with income are perforce recognized, and as individual preferences significantly change so does the scale of preferences in the economy as a whole. Therefore production decisions in capitalism are necessarily decentralized or diffused. Hence the problem of "planning under freedom" is insoluble, if by planning is meant the centralized adaptation of economic resources to a scale of preferences. If the hierarchy of wants

is planned there is no freedom; if freedom exists there can be no central plan.

SOME ASPECTS OF SOCIALISM IN RUSSIA

Quantity and Quality of Products

From time to time since about 1930, the Soviet press has published complaints from consumers about the quality of goods. This "socialist self-criticism" points to one feature of a socialist economy, namely, its probable tendency to stress quantity over quality. If the manager of a shoe factory is under pressure to increase output according to the plan, he will of course strive to turn out as many *pairs* of shoes as possible, especially if failure to do so might impugn his loyalty. Under capitalism, on the other hand, it is the *value* of output which counts and not mere physical quantity.

Unequal Incomes

Incomes in the Soviet Union are unequal, varying somewhat according to skill and efficiency. This is not a contradiction with the principles of socialism as Marxists use that term, for Marx and Lenin explicitly wrote that incomes would continue to be unequal under the socialism they proposed. The incomes of Russian authors, actors, and other artists are often very high, as much as hundreds of thousands of roubles a year. Manual workers are on piece-rate systems wherever possible (to a much greater extent in fact than in the United States), even workers on collective farms receive highly differentiated incomes. Managerial groups apparently earn much higher incomes than ordinary workers, and plant managers commonly receive bonuses if they lower costs or increase output beyond the assigned norms. The differentiation of money incomes in Russia does not, however, have the same effect as in the United States. The Russian income tax is comparatively low. In addition persons in key positions apparently also have prior rights (beyond those conferred by their money earnings) to acquire scarce consumer goods, such as large apartments, automobiles, country estates, and the services of household employees (servants).

Differential incomes in Russia emphasize the importance attached to incentives and penalties in the system. The "speed-up" in Russian factories is rewarded by extra pay, larger rations, and in conspicuous cases by medals. On the other hand, absenteeism from work and carelessness on the job might bring the wrongdoer banishment to a labor battalion in the Arctic

forests. Likewise, should a plant manager fail to meet his production quota he might be reduced in status or worse (there are reported instances of managers being shot for this reason). It should be recognized that the Russian temperament may impel this sharp juxtaposition of rewards and punishments, and also that the poverty of the country and its haste to expand may be accountable. Indeed, the Western world, in its ascent from the poverty of medieval times, previously punished and rewarded more fully than it does today. In the past the profits of the successful were large and also untaxed, but today they are heavily taxed; then, the "work-shy" were often thrown into a workhouse or transported to far-off colonies, but at present the worst that government (at least the American) may do is deny unemployment compensation benefits. As the Western world grew in wealth and increasingly accepted the spirit of humanitarianism in public affairs, penalties were moderated and rewards reduced. Whether Russia will follow suit in the course of time is a matter only of speculation. At present, however, it relies on liberal incentives and harsh penalties as planned techniques to hasten economic expansion.

Prices

The planning system in the Soviet Union uses the mechanism of prices and a money system including extensive banking facilities. It is not clear just what role prices play in the operation of the plan, for in this respect, as in others, detailed analysis has not been made available to foreign observers. It would seem that at the highest level, planning proceeds in physical terms and that money and prices seem only to provide a common denominator. Unlike the capitalist method whereby guidance of the productive system is largely based on the market (prices), guidance in Russia is according to plan, drawn up in physical units rather than values (prices). The individual producing units are responsible for minimizing their costs and maximizing output, although the measure of output may often be centered on quantity alone. The producing units earn profits, and these have more than a bookkeeping similarity to the profits under capitalism. For profits in Russia are used both as a method of control—high or low profits being taken as at least initial evidence of greater or lesser efficiency—and as a device for obtaining funds for fiscal purposes, for the profits accrue to government.

Economic Fluctuations

There is no business cycle in the ordinary sense in the Soviet Union. This does not mean, however, that production flows smoothly in a steady pattern of increase. In 1937, which was a year of extensive political trials, there seems to have been an actual decline in production. It may be that a socialist economy develops a rhythm peculiar to itself, just as the business cycle is peculiar to capitalism, and just as the alteration of feast and famine characterized the precapitalist societies. In any case it is too soon to judge whether or not production will continue to vary in the Soviet Union as in the past. Russia is by no means unaffected by prosperity and depression in the rest of the world, being an exporter (primarily of raw materials) and an importer (principally of machinery). The prices of raw materials and of machinery behave differently over the course of a business cycle, the former having a tendency to rise and fall much more than the latter. Hence, Russia is worse off than otherwise when the rest of the world is in a depression, because more raw materials have to be exported to acquire a given volume of machinery imports than in a "normal" period. The reverse holds true in periods of world prosperity.

Unemployment

"There is no unemployment in Russia." At least, there have been no figures published since the early 1930's which would indicate its existence. It is difficult to believe that there can be literally no unemployment, for any modern economy will constantly exhibit some frictional unemployment. Assuming that none exists, it should be noted that while the abolition of large-scale unemployment is distinctly an achievement it is dependent in part on the compulsion to work. A surplus of labor in one area is not allowed to be even temporarily unemployed; if wage incentives fail to induce migration to the point at which labor is needed, compulsion is used.

These various aspects of the Russian economy in operation inhere partly in the nature of socialism and partly in the peculiar context of Russian socialism. To operate effectively socialism requires the centralization of economic decisions together with the determination of social goals—or a hierarchy of wants, or a scale of preferences—wholly or largely by central authority. Selection of goals to be striven for is inevitable because the scarcity of resources makes it impossible to satisfy all wants; thus the subordination of individual preferences is unavoidable, just as is the imposi-

tion of authority on the individual's own efforts and time. As realists and direct actionists the Soviet leaders go to the point immediately; the goal is established, a plan is devised, and the scarce means at hand are relentlessly mobilized.

BRITISH SOCIALISM

With the victory of the Labour party in 1945, Great Britain embarked upon a program of partial socialization which had been proposed for many years by the party. The program is partial because the British socialists limit their program to certain basic industries such as coal, iron and steel, and transportation. Enterprise otherwise remains largely in private hands. At the end of two years, most of their socialization program had been adopted; the transport and coal industries had been seized, the Bank of England nationalized, and other reforms instituted. The first two years of operation has raised important questions about British socialization.

In the first place, the mere transfer of ownership from private to public possession does not obviate the necessity of facing economic realities. Some "sick industries," like the coal industry, can only be restored to high-level production by extensive reorganization, mechanization, and hard work and enterprise. A transfer of ownership to public authority as such does not assure such changes and efforts; indeed, socialization may intensify some of the difficulties if it encourages the hope that more will be produced with less effort. Not only in the socialized industries but throughout the British economy the necessity to produce increased amounts of national output is of utmost importance, for Britain must maintain large exports to sustain its levels of material well-being. To contribute to this end, the socialized industries, and especially the coal industry, must produce more, not less, than before. Socialization per se neither assures this, nor dispels the need.

The British socialist program is being operated within the framework of a government of limited powers—a democracy. Conceivably, it may be impossible within the scope of present powers for the government to employ sufficient force to manage the economy effectively. Moreover, reared in the democratic traditions of Britain, the socialist ministers may shrink from acquiring or seizing the power needed for effectiveness. The added powers granted the government in 1947 probably fall short in this respect. Furthermore, partial (but majority) political support tends to weaken the goals of a socialist program, for even the majority party's political actions must be qualified by the pressure of powerful groups, notably the unions,

among its constituents. The general goal of policy includes the rapid expansion of industrial capital to restore Britain as a great industrial power. This, however, requires a reduction in Britain's propensity to consume, by decreasing money wages, or by stabilizing them in the face of a rise in consumer prices, or by permitting them to rise less rapidly than prices. Imports of consumer goods also need to be curtailed. In short, what the British call "austerity" will have to be endured for a long time to accomplish the goal. The party's political support, however, comes from the groups who have to put up with most of the "austerity." Each group might argue that a little more income and consumption for it would not defer attainment of the goal; the combined effect of individual or group behavior consequently might make attainment impossible. If the government persists in its goal it may lose political support.

Faced with the possibility of losing political support if it continues with its general plan, or retaining political support at the sacrifice of its plan, a third course may present itself: to retain power and impose its goal at the expense of democratic traditions. If this third course appears improbable, there remains the other two choices, neither of which can be viewed happily by a politician in power.

The political problems and economic difficulties mentioned here are readily evident in the socialist program in practice. It is possibly because of these troubles that the British socialists are more inclined to "plan" than to execute a policy with vigor. Indeed, it is easier to plan than to execute a policy, for execution is rendered difficult when the goals are obscured—as they are in British socialism. Partial socialization implies either conflicts or obscure goals; in either case the effective mobilization of economic resources is impossible, for, as shown above, the ends of collectivist action must be clear if the means of attainment are to be effectively used. With quasi socialism (partial socialism), Britain will probably muddle through its economic difficulties, with no more, and possibly with less, success than the lethargic quasi capitalism characterizing its economy in recent decades.

SOCIALISM AND THE AMERICAN ECONOMY

The American economic system faces the basic economic problem confronted by all economic systems: a scarcity of economic resources in relation to the wants to be satisfied. The fundamental economic differences between American capitalism and a socialist system are to be found in (1) the mode of formulating the hierarchy of wants or scale of preferences of

the economy, and (2) the method of organizing scarce resources in a way appropriate to the satisfaction of these wants.

In capitalism, wants, for the most part, express individual or family preferences. What is wanted is influenced by many things, chief of which is production itself. But the goods made available in turn are governed by producers' expectations of what consumers will want. Wants and production therefore interact. The individual enjoys a wide range of discretion in forming his preferences; within the scope of his income he is free to choose from a great variety of goods and services the things he wants. He is free to employ resources (indirectly by buying what he wants and can afford) for his own personal ends. This may be termed *consumer autonomy*, an aspect of individualism which is frequently neglected and without which individualism in other spheres of the economy would not exist.

In contrast, the socialist economy centrally establishes the scale of preferences or hierarchy of wants which will direct the employment of resources. Wants regarded as inessential or undesirable by socialist planners will not be satisfied, simply because resources will not be devoted to goods and services aimed at their satisfaction. To be sure, the scale of preferences permitted by the planners or determined by them might be agreeable to the majority of consumers. But all consumers might be denied the satisfaction of some wants, and some might be denied the satisfaction of many wants which they prefer to those that are satisfied. In short, individual preferences are subordinated to the socialist planners' conception of "social wants"—presumably those thought by the planners to be good for people and the community or state.

In a capitalist system, the consumer or individual autonomy in forming preferences leads to the utmost diversity and unpredictable changes in the hierarchy of wants, and consequently in the use of economic resources. Autonomy in the formation of preferences presupposes a large measure of autonomy in the employment of resources, i.e., the institution of private property and the right to dispose of one's time, effort, and property in accordance with private interest. Given autonomy of preference and of employment of resources, the allocation or use of economic resources will be governed chiefly by the markets, wherein individual preferences are manifested.

This allocation for the most part is necessarily unplanned—but by no means does this mean that it is chaotic. Free choice by consumers means in effect the employment of resources by them (via intermediary producers); each purchase is a purchase of a given combination of human effort, capital, and natural resources required to produce the object bought. A cen-

trally planned use of economic resources cannot be in effect when free consumer choice is allowed to influence production. If the employment of resources is planned, then the scale of preference must be planned. In this way, capitalism and socialism in their pure forms are poles apart. The contrast is often drawn on the basis of legal ownership of resources—private versus public ownership; or on motives—private interest versus public interest. These contrasts, however, are essentially derived from the matter of choice—individual autonomy in the formation of preferences versus the social (or state) determination of the hierarchy of wants.

Capitalism in the United States is by no means unalloyed. There exist here both social goals and the planned use of economic resources to attain them. A large part of total governmental expenditures is devoted to such public purposes as highway construction, public education, public health, public housing, and social security. Each represents a part of what may be termed the *social policy* or objectives of government. Compared with a socialist state, however, these social goals require for their attainment only a small fraction of the nation's resources. The nearest approach of the American economy to a likeness with socialism has been during war, especially World War II. In that conflict the goal was military victory. The hierarchy of wants in the economy, and therefore the allocation of resources, was dominated by this goal.

The social policy of a government is in itself revealing of the contrast between individual autonomy and state autonomy in the formation of preferences. In some instances, for example highway construction or public education, individual and social goals for the most part coincide. Highway construction commands almost general approbation from all groups in the economy, perhaps because they are from a standpoint of use truly "public." Practically everyone uses the highways and there is no cleavage between income groups in this respect. The poor in their antiquated cars travel the same roads as the rich in their expensive cars. Only by public undertakings could their want for traveling convenience be satisfied. Public education and public health command general approval, but perhaps not so unqualifiedly as highway policy, for the direct benefits are not so evenly spread among all groups in the system, nor are they so essential to some as to others. Other social objectives raise controversial issues. Public housing, for example, benefits directly only a small part of the population although its indirect benefits may be more general. The wants of a relatively small number may be satisfied in accordance with public policy but possibly at the expense (through increased taxation) of the wants of others. Socialized medicine is similar in this respect. Its cost to some would reduce

their capacity to satisfy wants; the wants of others would be satisfied more fully than otherwise. Regulating want satisfaction by means of government therefore results in subordinating the satisfaction of wants by some, and possibly by many. Whose wants, and which wants, should be satisfied? In the United States individual preferences still dominate the economic processes, although the establishment of social goals through government has grown in importance.

In the experience of a capitalist economy it is found that the problem of subordinating individual and group interest proves troublesome even during wartime, when there necessarily exists an integrated economy dedicated to a widely accepted social objective. A few illustrations indicate the problems involved. In gasoline rationing, graduations of essentiality were established and differential rations were granted on the basis of them. Rationing officials were called upon to determine essentiality—in effect, the extent to which automobile-driving wants were to be satisfied. But in spite of the war emergency, complaints and evasions were widespread. Again, in the allocation (rationing) of scarce materials to industry, the basis was essentiality to the total war needs. Complaints and evasions were common because individual (and group) conceptions of essentiality often varied sharply with those of the custodians of public interest. Price control raised similar problems of group dissidence from social objectives as defined by governmental officials. Political pressure from the farm group succeeded in keeping farm prices from control for a time, and when controls were finally extended to farm products their inflated prices were actually maintained in the legislation authorizing controls. The problems involved in subordinating individual and group ends or purposes to the national goal are obviously many.

In the national emergency of war compulsion is necessary—and feasible. In the absence of an emergency so great as war, the establishment of *over-riding* social goals to be attained by collective action is inconsistent with both the ideals and the practices of a free society. These ideals and practices underlie the processes in the modern American economy: the freely expressed scale of preferences, and the free mobilization and use of resources.

SELECTED REFERENCES

PART I THE SCOPE OF ECONOMICS

CHAPTERS 1 AND 2

Knight, F. H., *The Ethics of Competition,* Harper, New York, 1935 (especially chs. 1 and 2)

Robbins, Lionel, *An Essay on the Nature and Significance of Economic Science,* Macmillan, New York, 2nd ed., 1935

"Economics," Encyclopedia of the Social Sciences, Macmillan, New York, 1931

Johnson, E. A. J., *Some Origins of the Modern Economic World,* Macmillan, New York, 1936

Tawney, R. H., *Religion and the Rise of Capitalism,* Harcourt, Brace, New York, 1926

PART II NATIONAL INCOME AND THE EMPLOYMENT OF RESOURCES

CHAPTERS 3-7

Dewhurst, J. Frederick, and Associates, *America's Needs and Resources,* Twentieth Century Fund, New York, 1947 (especially chs. 1-3, 22-25)

Shoup, Carl, *Principles of National Income Analysis,* Houghton Mifflin, Boston, 1947

National Income, supplement to *The Survey of Current Business,* July 1947, U.S. Department of Commerce, Washington, D. C.

Keynes, J. M., *The General Theory of Employment, Interest and Money,* Harcourt, Brace, New York, 1936

Robinson, Joan, *Introduction to the Theory of Employment,* Macmillan, New York, 1937

Morgan, Theodore, *Income and Employment,* Prentice-Hall, New York, 1947

Haberler, Gottfried, *Prosperity and Depression,* 1941 ed., United Nations, Lake Success, New York

Schumpeter, Joseph A., *Capitalism, Socialism, and Democracy,* Harper, New York, 1942 (especially chs. 5-14—capitalist disintegration view)

Hayes, H. Gordon, *Spending, Saving, and Employment,* Knopf, New York, 1945 (mature economy view)

Terborgh, George, *The Bogey of Economic Maturity* (1945), and *An Appraisal of the Fatalistic View of Capitalism* (1944), The Machinery and Allied Products Institute, Chicago, Illinois (expansive economy view)

PART III THE MONETARY PROCESS

CHAPTERS 8-11

Curtis, Myra, and Townshend, Hugh, *Modern Money,* Harcourt, Brace, New York, 1938

Halm, George N., *Monetary Theory,* Blakiston, Philadelphia, 1942

Fisher, Irving, *Purchasing Power of Money,* Macmillan, New York, 1911

The Federal Reserve System; Its Purposes and Functions, Board of Governors of the Federal Reserve System, Washington, D. C., 1947

Robertson, D. H., *Money,* Harcourt, Brace, New York, 1929
Banking Studies, Board of Governors of the Federal Reserve System, Washington, D. C., 1941

PART IV PRICES AND THE ALLOCATION OF RESOURCES

CHAPTERS 12-16

Wootton, Barbara, *Plan or No Plan,* Farrar and Rinehart, New York, 1935
Stigler, George, *The Theory of Price,* Macmillan, New York, 1946
Marshall, Alfred, *Principles of Economics,* 8th ed., Macmillan, New York, 1920 (especially Books III and V)
Chamberlin, Edward, *The Theory of Monopolistic Competition,* 5th ed., Harvard Univ. Press, 1946
Robinson, Joan, *The Economics of Imperfect Competition,* Macmillan, London, 1933
Hicks, J. R., *Value and Capital,* Oxford Univ. Press, London, 1939
T N E C Monograph No. 1, *Price Behavior and Business Policy,* Government Printing Office, Washington, D. C., 1940

CHAPTERS 17-21

Marshall, *op. cit.* (especially Book VI)
Clark, J. M., "Distribution," Encyclopedia of the Social Sciences, V, Macmillan, New York, 1931
Readings in the Theory of Income Distribution, Blakiston, Philadelphia, 1946
Stigler, *op. cit.*
Douglas, Paul H., *The Theory of Wages,* Macmillan, New York, 1934
Hicks, J. R., *The Theory of Wages,* Macmillan, New York, 1932
Robinson, Joan, *Essays in the Theory of Employment,* Macmillan, New York, 1937
Robinson, Joan, *The Economics of Imperfect Competition,* Macmillan, London, 1933
Fisher, Irving, *The Theory of Interest,* Macmillan, New York, 1930
Keynes, *op. cit.*
Knight, F. H., *Risk, Uncertainty, and Profit,* Houghton Mifflin, Boston, 1921

CHAPTERS 22-23

Mills, F. C., *The Behavior of Prices,* 1927, and *Prices in Recession and Recovery,* 1938, National Bureau of Economic Research, New York
Dewhurst, *op. cit.*
Fisher, Irving, *The Making of Index Numbers,* Houghton Mifflin, Boston, 1922
Black, J. D., *Parity, Parity, Parity,* Harvard Univ. Press, Cambridge, 1942
Office of Price Administration, *A Manual of Price Control,* Government Printing Office, Washington, D. C., 1943
Harris, S. E., *Inflation and the American Economy,* 1945, and *Price and Related Controls in the United States,* 1945, McGraw-Hill, New York
T N E C Monograph No. 32, *Economic Standards of Government Price Control,* Government Printing Office, Washington, D. C., 1941

PART V MANAGEMENT AND CONTROL OF ECONOMIC RESOURCES

CHAPTERS 24-30

Berle, A. A., and Means, G. C., *The Modern Corporation and Private Property,* Macmillan, New York, 1933
Dewing, A. S., *The Financial Policy of Corporations,* 4th ed., Ronald, New York, 1941
Gordon, Robert A., *Business Leadership in the Large Corporation,* The Brookings Institution, Washington, D. C., 1945
Twentieth Century Fund, *Big Business, Its Growth and Place,* New York, 1937
Millis, H. A., and Montgomery, R. E., *The Economics of Labor,* 3 v., McGraw-Hill, New York, 1938-45

Slichter, Sumner, *Union Policies and Industrial Management*, The Brookings Institution, Washington, D. C., 1941
Readings in the Social Control of Industry, Blakiston, Philadelphia, 1942
Lyon, L. S., Watkins, M. W., and Abramson, V., *Government and Economic Life*, The Brookings Institution, Washington, D. C., 1939

PART VI GOVERNMENT FINANCE

CHAPTERS 31-33

Allen, E. D., and Brownlee, O. H., *Economics of Public Finance*, Prentice-Hall, New York, 1947
Groves, Harold M., *Financing Government*, rev. ed., Henry Holt, New York, 1945
Groves, Harold M., *Postwar Taxation and Economic Progress*, McGraw-Hill, New York, 1946
Simons, Henry C., *Personal Income Taxation*, Univ. of Chicago Press, Chicago, 1938

PART VII INTERNATIONAL ECONOMICS

CHAPTERS 34-36

Heilperin, Maurice A., *The Trade of Nations*, Knopf, New York, 1946
A Review of Export and Import Procedure, Guaranty Trust Co., New York, 1946
Patterson, E. M., *An Introduction to World Economics*, Macmillan, New York, 1947
Enke, Stephen, and Salera, Virgil, *International Economics*, Prentice-Hall, New York, 1947
Hoover, Calvin B., *International Trade and Domestic Employment*, McGraw-Hill, New York, 1945
Dewhurst, *op. cit.*, ch. 21

PART VIII PERSONAL DISTRIBUTION AND USES OF INCOME

CHAPTERS 37-40

Leven, M., Moulton, H. G., and Warburton, C., *America's Capacity to Consume*, The Brookings Institution, Washington, D. C., 1934
Consumer Incomes in the United States, National Resources Committee, Government Printing Office, Washington, D. C., 1938
Consumer Expenditures in the United States, National Resources Board, Government Printing Office, Washington, D. C., 1939
Dewhurst, *op. cit.*, chs. 5-15
Readings in the Theory of Income Distribution, Blakiston, Philadelphia, 1946
Burns, Arthur E., and Williams, Edward A., *Federal Work, Security and Relief Programs*, Government Printing Office, Washington, D. C., 1941

PART IX TYPES OF ECONOMIC SYSTEMS

CHAPTERS 41-42

Schumpeter, *op. cit.*
Hayek, F. A., *The Road to Serfdom*, Univ. of Chicago Press, Chicago, 1945
Wootton, Barbara, *Freedom under Planning*, Univ. of North Carolina Press, Chapel Hill, 1945
Baykov, A., *The Development of the Soviet Economic System*, Cambridge, England, 1946
Lange, Oscar, and Taylor, F. M., *On the Economic Theory of Socialism*, Univ. of Minnesota Press, Minneapolis, 1938
Knight, F. H., *Freedom and Reform*, Harper, New York, 1947

INDEX

ability to pay principle of taxation, 711-16, 820
absolute advantage, 740
abundance, meaning of, 9, 913
acceleration, principle of, 137, 140, 147
accumulation of capital, 55, 429, 430, 433; *see* capital
acid test, in financial analysis, 577
ad valorem duty, 788; *see* tariffs
adverse clearing balance, 227, 246; *see* clearing and collection
advertising, 291, 339
advertising competition, 354, 360-62
agents of production defined, 394; *see* factors of production
aggregate demand, 16, 17, 91 *ff.*, 107-09
  contrasted with particular demand, 288
  as determinant of output, 127 *ff.*
  in postwar period, 150
  relation to money supply, 253-55
aggregate expenditures, 91; *see* aggregate demand
aggregate supply, 107-09
  contrasted with particular supply, 288
  see gross national output
Agricultural Adjustment Act of 1933, 491, 502, 519, 661, 662
agricultural price control, and exports, 503, 504
agricultural prices, 491-94, 500-06; *see* parity prices
agriculture, governmental policy, 661, 662
air transportation, regulation of, 658, 659
allocation of consumer expenditures, *see* consumer expenditures
allocation of resources, *see* resources, allocation of
Altman, Oscar L., 432 *n.*
amalgamated union, 617
American agriculture, character of, 450
American economy, growth of, 52-56
American Federation of Labor, 611, 614 *ff.*
anti-monopoly movement, 587, 588; *see* antitrust laws; *see* Sherman Act *and* Clayton Act
antitrust laws, 653-55; *see* Sherman Act *and* Clayton Act
antitrust laws in labor disputes, 632
antiunion activities of employees, 633, 634
arbitration
  compulsory, 639
  voluntary, 638
Aristotle, 18, 109
assets of a corporation, 574, 575; *see* balance sheet *and* financial statements
assets of commercial banks, 225
average costs, 319-25, 399
average fixed costs, 319, 322
average product, value of, 397, 398; *see* diminishing returns
average revenue, 348, 349
average total costs, 319-25
average variable costs, 319-25

balance of international payments, 753-66
balance sheet of corporations, 574, 575
bank examination, 224
bank failures, 216, 219-21
bank credit, 227, 228; *see* demand deposits
bank loans, 225, 227
bank reserves, and expansion of deposits, 229 *ff.*
bank reserves, 225, 226, 233, 246
  partial, 269; *see* legal reserve requirements, legal reserves, *and* excess reserves
Bank of England, 264-66
Bank of the United States, first, 207
Bank of the United States, second, 207
bankers' banks, 236
Banking Act of 1935, 241, 268
banking reform, 263, 264
bankruptcy, 577-79
bargaining theory of wages, 409
barter, 180, 181

basing points, 368, 499, 596
multiple, 369
benefit principle of taxation, 711, 712
Berle, A. A., 554 *n.*
Beveridge Report, 875
Beveridge, Sir William, 78
bilateral trade, 750; *see* international trade
bimetallic standard, 189, 190, 259
birth rates, 77, 78 *n.*, 79 *n.*
Bituminous Coal Acts, 507
bituminous coal prices, 506, 507
black list, 663
black markets, 279, 518
Bland-Allison Act of 1878, 260
blind, aid to, 877
"blue-sky laws," 580, 581
Board of Governors of the Federal Reserve System, 212-14, 241, 247, 268, 583
function of, 218, 219
bonds, *see* corporation bonds *and* government bonds
bond yields, 490, 491
book value of corporation, 572 *n.*
boycotts, by unions, 624, 625
branch banking, 222, 223
Bretton Woods Conference, 775, 794; *see* International Monetary Fund
Brewster, Sir Francis, 40
British socialism, 928, 929; *see* socialism
Brookings Institution, 823, 824 *n.*
Budget and Accounting Act of 1921, 689, 690
Budget, Bureau of, 689, 690
budget, federal, balancing of, 723 *ff.*; *see* deficit financing
building and loan associations, 204, 205
bureaucracy
in business, 533, 534
in government, 538-40
in labor organizations, 536-38
Burns, Arthur E., 724 *n.*, 726 *n.*
Burns, A. R., 368 *n.*
business cycles, 127, 485; *see* chapter 6
in Russia, 927
business failures, 577-79
business capital, sources of, 560, 562-69; *see* capital
business combination, 590-92
business financing, methods of, 560; *see* capital *and* corporation
business finance, 559 *ff.*; *see* corporation
business profits, 460 *ff.*; *see* profits
business reorganization, 577-79
business organizations, 541 *ff.*
business savings, 119-21; *see* savings
business taxes, 693, 699, 700
indirect, 88, 89, 119-21, 123
state, 704, 705
business unionism, 621
bylaws of corporations, 547

Cagle, C. E., 223 *n.*
call loan, 441
call-money rates, 489
capital
accumulation of, 55, 429, 430, 433
and business enterprise, 559 *ff.*
circulating, 560, 561
defined, 10, 69, 70, 559
depreciation of, 72, 73
exports of, 780-85
facilities, 65, 70, 71
formation of, 146, 429 *ff.*
goods, and national output, 94, 119
goods, services of defined, 392
growth of, 51, 52
imports, 780-85
items in balance of payments, 758
marginal efficiency of, 128 *n.*
marginal product of, 434, 438, 439
and money, 70 *n.*, 392
and national output, 146
operating, 112
replacement of, 72, 88
resources, 65
relation to saving, 10, 429 *ff.*
capital gains taxation, 698, 699
capitalism
defined, 274, 893, 894
development of, 166 *ff.*
and democracy, 280, 914, 915
disintegration thesis, 165-73
regulated, 50
contrasted with socialism, 929-32
socialist critique of, 897-99
capitalist economy, 60
Carlyle, Thomas, 896, 915
cartels, international, 800-02
central banking policy, 264, 265; *see* Federal Reserve System
central banks, 207, 759, 760
functions of, 212
central reserve city banks, 210, 214

certificates of indebtedness, Treasury, 728; *see* national debt
chain banking, 223
chain stores, taxation of, 705
checks, 185, 186
  clearing and collection of, 216-18; *see* demand deposits *and* money
check-off, 629
Church, the, 30, 31, 33, 34, 51, 168
circulating capital, 560, 561
Civil Aeronautics Administration, 658
Civil Aeronautics Board, 658, 659
Clark, Colin, 81 *n.*
Clark, J. M., 336 *n.*
classes, social, relation to productive services, 386, 387
"class struggle," 900-01, 907
classified property tax, 703
Clayton Act of 1914, 363, 406, 499, 594-96, 600, 632, 653
clearing agreement, 797, 798
clearing and collection of checks, 216-18
clearing balance, *see* clearing and collection of checks
clerical workers, 76
closed shop, 536
cobweb theorem, 344, 345
"codes of fair competition," 520, 521; *see* National Recovery Administration
coins, 184, 197, 200
collective bargaining, 499, 626-30
collectivism, 913; *see* communism, fascism, socialism
  relation to foreign trade, 743, 750
Colm, Gerhard, 693 *n.*
collateral trust bonds, 567, 568; *see* corporation
combined duty, 788; *see* tariffs
Commerce, Department of, 85, 88 *n.*, 92, 95, 97, 99, 103, 145, 147-49
commercial banks, 203 *ff.*, 436, 437
  assets of, 225
  defined, 206
  functions of, 203, 206
commercial enterprises of government, 687, 688
commercial paper, 198, 216, 569; *see* rediscounting *and* discounts
Commercial Revolution, 36
Committee for Economic Development, 554
commodity agreements, international, 799, 800
common stock, 550, 563; *see* corporation
communism, 668, 906, 907, 911, 913-15; *see* Soviet Union *and* Marx
company union, 616
comparative advantage, principle of, 740-43, 787, 788, 805; *see* international trade
compensated dollar, 257
compensation agreement, 797
compensatory spending, 725; *see* deficit financing
competition, 335 *ff.*, 540
  defined, 336
  function of, 531
  see monopolistic competition, oligopoly, *and* pure competition
competitive economy, 274
complementary goods, 311
  price relations between, 374, 375
Comptroller of the Currency, 208, 224
Comptroller General, 689, 690
compulsory arbitration, 639
compulsory investigation, 638, 639
concentration, 495-98
  of corporate control, 640
  of corporate wealth, 554, 555
  of industry, 585 *ff.*
  techniques of, 593
conciliation, 638
Congress of Industrial Organizations, 616 *ff.*
conservation of natural resources, 68, 69, 662, 663
constant-cost industries, 331
Constitution of United States, 648-51
consumer co-operatives, 557
consumer, the, *see* individual demand
  preferences, 296
  problems of, 289-92
consumer expenditures
  and allocation of resources, 845, 846
  distribution of, 289
  and income distribution, 842-45
  influence on investment, 113-19
  defined, 9
  and national output, 94
consumer goods, *see* consumption
  defined, 9
  and national output, 94

consumer goods (Cont.)
distinguished from productive services, 383
consumer price index, 481
consumption, 23, 24
consequences of, changes in, 135 *ff.*
induced increase in, 135 *ff.*
under mercantilism, 42, 43
in medieval economy, 31, 32
and national output, 119
ratio, 100
theory of, 292-301
see demand *and* consumer expenditures
consumption taxes, 693, 694, 701, 702, 705-07, 714, 715
control, private and public, 666-68
convertibility of money, 189-91
cooperatives, 556-59
Corn Laws, 447
corporate state, 912
corporation, 542 *ff.*
book value of, 572 *n.*
bonds, 550, 565-68
charter, 562
defined, 545
management of, 547, 548
ownership of, 547, 550
stock, 550
corporation income tax, shifting of, 720, 721
correspondent banks, 209
costs
average, 319-25, 399
average fixed, 319, 322
average total, 319-25
average variable, 319-25
business costs, classification of, 385
direct, 385
of a firm, 317-25
fixed, 318-22
human, 11, 12
indirect, 385
long-run, *see* long-run supply schedule
marginal, 319-25, 399
opportunity, 419, 460, 461
overhead, 318-22, 385, 386
prospective, 111, 112, 114-16
real, 11, 12
and scarcity, 12
short-run, 318
variable, 318-22
cost of living, 478, 481
and wages, 423
cost-of-living index, 477, 481, 489
country banks, 210, 214
Council of Economic Advisers, 727
craft guild, 33, 36
craft union, 613, 617
credit, *see* bank credit, bank loans, *and* demand deposits
credit items in balance of payments, 753, 754, 757, 761-63
creditor nation, 780
"crime of '73," 190
Crowder, W. F., 496 *n.*
cumulative preferred stock, 564; *see* corporation
currency, 184
in circulation, effect on member bank balances, 236-40, 247
in circulation, changes in World War II, 201
system of United States, 197-201
see money *and* coins
currency depreciation, 770-72; *see* devaluation *and* gold dollar
current income, 102
cycles, *see* business cycles
cyclical unemployment, 869, 870

Daugherty, C. R., 830 *n.*
death rates, 77
death taxes, 700, 701; *see* estate taxes *and* inheritance taxes
debenture bonds, 566, 567; *see* corporation
debit items in balance of payments, 753-55, 757, 763-65; *see* balance of international payments
debtor nation, 782
decisions, economic, 24-28, 106-10, 527 *ff.*
decreasing-cost industries, 331
deficit financing, 724-27
effects on banking system, 242-44
deflation, 245, 255, 256, 519
demand (for consumer goods)
changes in, 303, 304
expected, 107, 113, 116
individual, 292-301
individual demand curve, 299
individual demand schedule, 298, 299
interrelations of, 310, 312
law of, 302

demand (Cont.)
market, changes in, 303, 304
market demand schedule, 301 *ff.*
particular and aggregate contrasted, 288
in short-run, 342-45
see elasticity of demand
demand for foreign exchange, 767-80
demand for money, 270
demand for productive services, 382 *ff.*
for capital, 431 *ff.*
for labor, 410-14
for land, 446 *ff.*
demand deposits, 225, 226
creation of, 243
control of, 262, 268-71
defined, 185
expansion of, 229 *ff.*
see money
democracy and capitalism, 280, 914, 915
dependent children, aid to, 877
deposit insurance, 219-21
deposits, *see* demand deposits, time deposits, money
depreciation, 72, 73, 87, 88, 91, 102, 119-21
depreciation of currency, 770-72; *see* devaluation *and* gold dollar
depreciation reserves, 109
depression unemployment, 83, 869, 870
derived demand for productive services, 383, 855
devaluation
effects on gold stocks, 258, 259
and prices, 257, 258
of United States dollar, 194, 195, 256, 258, 771
deterrents to investment, 159, 163, 164; *see* expansive economy thesis
"dialectical materialism," 904
"dictatorship of the proletariat," 906, 914, 915, 922
differential rents, 450-53; *see* rent
diminishing marginal productivity, 857; *see also* diminishing returns
diminishing marginal utility, 293-95, 299-301
diminishing returns, law of, 397-401, 407, 411, 434, 448, 452, 453, 458
direct costs, 385
discounts and advances, 237
disinvestment, 134
distribution of income, *see* income distribution
dividends, 90, 563, 569; *see* corporation
division of labor, 47; *see* specialization
dollar
balances, 760
defined, 186, 187
devaluation of, 194, 195, 256, 771
exchange, 752, 760
value of, 249, 256
Douglas, Paul H., 488 *n.*
drafts, 761; *see* foreign exchange, financing of exports, financing of imports
due process clause, 649, 650
dumping, in foreign trade, 792
duopoly
defined, 363
prices, 364
durable goods, 134
Durand, David, 490 *n.*
dutiable imports, 788; *see* tariffs
duty, tariff, *see* tariffs

earned income, 97
earnings of labor, 487-89; *see* wages
economic agents, 521, 522; *see* factors of production
economic goods, defined, 8, 9
economic interpretation of history, 904, 905
economics, defined, 2, 3, 11
economizing, necessity of, 917, 918
economy, the mixed, 50
effective demand, 298; *see* demand
elasticity of demand, 304-08, 351; *see* demand *and* income elasticity of demand
elasticity of demand for labor, 413
elasticity of supply, 316, 332
elasticity, unit, 305
employers' liability laws, 879, 880
Engels, Friedrich, 900, 917
entrepreneur, 387, 393, 463, 464
entrepreneurial services, defined, 393
equation of exchange, 250-52, 257; *see* quantity theory of money
equilibrium amount, 341
equilibrium price, 340
equipment trust bond, 567; *see* corporation
equity capital, 562; *see* corporation
equity structure of corporation, 569-72

estate taxes, 700
"ever-normal granary," 502
excess profits tax, 699, 700
excess reserves, 214, 226, 233, 269; *see* bank reserves
Exchange Equalization Account, 773, 774
excise taxation, *see* consumption taxes
expansive economy thesis, 158-65
expectations
  short-term, 113-16, 121
  long-term, 116-18
experience rating, in unemployment compensation, 883, 884
explicit interest, 428
explicit profits, 460
explicit rent, 448, 449
explicit wage, 405
exploitation, Marx's concept of, 901-04
export transactions, financing of, 761-63
exports
  in balance of payments, 754, 757
  capital, 780-85
  world, 744, 745
extensive margin of cultivation, 451

Fabian socialism, 909, 910
factor costs, 89, 388, 389
factors of production, 64, 381, 387, 394
Fair Labor Standards Act, 510, 611, 861
fair-trade laws, 654
Farm Credit Administration, 206
farm income, 861
farm marketing co-operatives, 558
farm prices, 491-94, 500-06; *see* parity price
farm proprietors as gainful workers, 76
fascism, 668, 912, 913
favorable balance of trade, 253, 749, 751, 756
  in mercantilism, 36
Federal Advisory Council, 218
Federal Bankruptcy Law, 578, 579
federal debt, composition of, 728; *see* national debt
Federal Deposit Insurance Corporation, 220, 221, 224
federal expenditures, 674, 675
Federal Farm Board, 501-03, 661
Federal Home Loan Bank Board, 205
Federal Land Banks, 206, 558, 661
Federal Loan Agency, 206
Federal Open-Market Committee, 218
Federal Power Commission, 659
Federal Reserve Act, 212, 265
Federal Reserve Agent, 213
Federal reserve bank notes, 197, 199
Federal reserve banks, 197, 258, 759, 776
  assets and liabilities, 234-37
  location, 212
  ownership, 213
Federal Reserve Board, *see* Board of Governors of the Federal Reserve System
Federal reserve districts, 212
Federal reserve notes, 197, 198, 201, 213, 226
Federal Reserve System, 212, 213, 442, 443; *see* Federal reserve banks *and* Board of Governors of the Federal Reserve System
federal savings and loan associations, 205
federal securities, ownership of, 728, 729; *see* national debt
Federal-state old age assistance, 889
federal tax system, 694 *ff.*
Federal Trade Commission, 596, 600-02, 654
feudal economy, 29, 30
fiat money, 193
Fifth Amendment, 649, 650
"finance capitalism," 906
finance companies, 205
financial institutions, types of, 204-06
financial ratios, 577; *see* balance sheet *and* financial statements
financial statements, 572-76
fiscal agents, Federal reserve banks as, 218
fiscal policy, 709 *ff.*; *see* deficit financing
fiscal monopolies, 664
Fisher, Irving, 250, 257
five-year plans in Russia, 917, 923-25
fixed capital, 560, 561
fixed costs, 318-22
flexibility of tax revenues, 722
flexible prices, 494
foreign bank balances and international financing, 759-66
foreign exchange
  control of, 770, 796-99
  financing of international transactions, 760, 761
  under gold standard, 767-72
  instability of, 769-76

foreign exchange (Cont.)
rate determination, 766 *ff.*
foreign investment, 780, 781-85
foreign trade
dependence of United States on, 737
and employment, 805-12
and national income, 805-12
of the United States, 746-50
foreign trade monopolies, governmental, 802, 803
Fourteenth Amendment, 649, 650
free and unlimited coinage, 189
free goods, 8, 9
free imports, 788, 789; *see* tariffs
free trade, 44, 785-88, 805
freight equalization, relation to prices, 369-71
frictional unemployment, 870, 871
Fund, *see* International Monetary Fund
full employment, 82, 151
functional distribution of income, 813

gainful workers, 75-77, 605-10
gasoline taxes, 705, 706
George, Henry, 909
General Accounting Office, 689, 690
General Maximum Price Regulation, 513
general strikes, 623
Geneva Conference on Trade and Employment of 1947, 794, 809
gift tax, 700
going-concern value of corporations, 570-72
gold
in balance of payments, 755-57
bullion, 196
coins, 194 *n.*, 196
dollar, 187, 189, 190, 194, 195
dollar, reduction in weight of, 256-58; *see* devaluation
export point, 769, 771
flow, and balance of trade, 751, 757-59
import point, 768, 771
imports, 269
nationalization of, 194, 195
price of, 194, 256-58
reserves, 191
shipping points, 769, 771
and silver, 188, 252
stock, 253
stock, effect on member bank balances, 236, 239, 240
gold certificates, 191, 194, 195, 197-99, 234, 247
Gold Reserve Act of 1934, 195
gold standard, 187-94, 260
abandonment of, 194, 769-72
international aspects, 192, 769-72
goods
complementary, 311, 374, 375
economic, 8, 9
see consumer goods *and* producer goods
government assurance of income, 865 *ff.*
government bonds, 441, 727, 728; *see* government securities *and* national debt *and* open market operations
government borrowing, 143, 687, 727-33
government control, areas of, 651-61
government expenditures, effects on national income, 131, 132, 142-44; *see* public expenditures
government finance, 669 *ff.*
government finance, monetary consequences of, 253, 254
government lending agencies, 206
government output and income payments, 123
government ownership, 664-66
government regulation, 538-40, 648 *ff.*
government securities, 235, 237, 242
types of, 727-29
see national debt
government services and national output, 119
government services as a component of national output, 94
greenbacks (United States notes), 190, 193, 199, 243, 727
Gresham's law, 190
gross national output, defined, 86, 87; *see* national output
gross national product, *see* gross national output *and* national output
gross savings, business, 108
group banking, 223
guaranteed bonds, 567; *see* corporation
guild socialism, 910
guilds, 33, 34, 859

Hansen, Alvin H., 153 *n.*, 488 *n.*
Hamilton, Alexander, 501
Hawley-Smoot tariff of 1930, 793
Heckscher, 40 *n.*
hedging, 470

Hegel, 904
Hicks, J. R., 69
"high-consumption" economy, 124, 141
Hobbes, Thomas, 171
holding company, 555, 593
  and corporate concentration, 549-53; *see* concentration
horizontal combination, 591
hours of work, 610, 611
Hudson's Bay Company, 546
Huntington, E., 75 *n.*
hyperinflation, 252, 476

immigration, 77-79, 607-10
  effect on wages, 416
implicit interest, 428
implicit profits, 460
implicit rent, 448, 449
implicit wages, 405
import quotas, 795, 796
import transactions, financing of, 763-65
imports, in balance of payments, 754, 755, 757
imports, world, 744, 745
imputed wages, 405
incentives and income, 820, 821
incidence of taxation, 717-22
income bond, 567; *see* corporation
income, disposable, 99
income, disposition of, 98-101
  relation to level of output, 831-33
income, distribution of, 21-23
  and capital accumulation, 55
  and consumer expenditures, 842-45
  and consumption, 833 *ff.*
  criteria of, 819-21
  determinants of, 847 *ff.*
  by income classes, 822-30
  in Middle Ages, 35
  and need, 848, 849
  and progressive taxation, 838, 840, 844
  and propensity to consume, 821
  under socialism, 920
  and spending-saving ratios, 814-16, 831 *ff.*
  see income inequality
income elasticity of demand, 309, 316, 334, 844
income equality, 848
income, inequality of, 822-31
  and ownership of wealth, 852-54
  persistence of, 830-31
income (Cont.)
  in Russia, 925
income, national, *see* national income
income payments, 90
  sources of, 120
income and productivity, 850, 851
income redistribution
  and national output, 818-21
  effects on spending and saving, 835, 836, 840, 843
income and services, 816, 817
income and social status, 849, 850
income statement, 572, 573
income tax, corporation, 699
income tax, individual, 694-99
  incidence of, 720
income taxes, effects on disposable income, 840, 841, 844
incomes, participating, 817-21, 851, 852
incomes, non-participating, 817-21, 851, 852, 873
inconvertible paper standard, 193
increasing-cost industries, 331
index numbers, 249, 476 *ff.*
Independent Treasury System, 208
indirect business taxes, relation to national output, 88, 89, 119-23
indirect costs, 385
individualism, 540
industrial accidents, 872
industrial banks, 205
industrial union, 613, 617
Industrial Workers of the World, 910, 911
inelastic demand, 305, 307, 308; *see* elasticity of demand
inelastic supply, 316; *see* elasticity of supply
inequality of incomes, *see* income, distribution of, *and* income, inequality of
infant industry argument, 789, 790; *see* tariffs
inflation, 255, 256
inheritance and concentration of wealth, 853, 854
inheritance taxes, 700, 701
injunctions in labor disputes, 631-33
instrumental goods, 10
insurance, 469
insurance of bank deposits, 219-21
intensive margin of cultivation, 451
interbank deposits, 210

interdependence in world trade, 745; *see* foreign trade
Interdistrict Settlement Fund, 217
interest, 392
  and capital, 428
  and capitalization, 443-45
  defined, 428
  function of, 444, 445
interest rates, 489-91
  structure of, 441-45
interlocking directorates, 595
International Bank for Reconstruction and Development, 784, 785, 811
international capital movements, 780-85
international cartels, 800-02
international commodity agreements, 504-06
international economic policy, 785 *ff.*
international finance, 753 *ff.*; *see* foreign exchange
International Monetary Fund, 775-80, 784, 785, 811
international payments, balance of, 753-66
international trade, 734 *ff.*; *see* foreign trade
  and specialization, 738-43
international trade balances as investment, 72 *n.*
International Trade Organization, 811
international unions, 612 *ff.*
interstate commerce
  control of, 649
  relation to control of monopoly, 598, 599
Interstate Commerce Act, 656
Interstate Commerce Commission, 364, 656, 657
interstate trade barriers, 655
investment
  changes in, 128-31, 134-39
  and consumer expenditure, 113-19
  decline during 1930's, 144 *ff.*
  deficiency of, 157, 159
  defined, 72
  expenditures, 17
  goods, 70
  gross, 72
  long-term, 112, 116-19
  net, 72, 105
  and productive capacity, 122
  and purchasing power, 120, 121
investment (Cont.)
  and savings, 17, 55, 121, 122
  short-term, 113-16
  and unemployment, 871
investment banks, 205, 206, 580-84
investment goods, 70
invisible hand, 45
irreducible minimum of unemployment, 871
isolated bank, 229-31

joint stock company, 546
joint supply, 332, 376
Jones, Martin V., 158 *n.*
jurisdictional disputes, 614, 620
just price, 34, 40

Keynes, J. M., 85, 153 *n.*, 172, 183, 725
King, W. I., 852, 853 *n.*
Knight, F. H., 4
Knights of Labor, 613, 625

labor
  defined, 390
  force, 414
  income, 97; *see* wages
  as a resource, 64, 65, 75-78
  scarcity of, 67, 74
  supply of, 75-79
labor legislation, effect on labor supply, 860
Labor-Management Relations Act of 1947, 636, 637
labor movement, 611 *ff.*
labor organization, 605 *ff.*
laborsaving and capital, 74
Labor Statistics, Bureau of, 477, 479, 481, 482
labor theory of value, 901, 907, 908
labor unions, 27, 164, 169, 173, 534-38, 605 *ff.*
Labour party, 928, 929
laissez faire, 44, 46, 48-50, 56, 647, 786, 915
land, 64-67
  supply characteristics of, 390
Landis, Paul H., 607 *n.*
large-scale production, economies of, 589, 590
law of demand, 302; *see* demand
law of diminishing returns, 448, 458; *see* diminishing returns

law of supply and demand, 287; *see* supply *and* demand
laws, natural, 41, 45
legal reserves
  Federal reserve banks, 234, 235
  member banks, 214, 240, 241, 268
  national banking system, 210, 211
  see bank reserves *and* demand deposits
legal tender, 198
Lend-Lease, 503
Lenin, 905-08, 914, 915, 921, 925
letter of credit, 761, 762
Leven, M., 823, 824 *n.*
leverage effect, 138; *see* acceleration, principle of, *and* multiplier
liabilities of a corporation, 574, 575
liabilities of commercial banks, 225, 226
liberalism, 42-51, 647-51
liquidity preference theory of interest, 435-40; *see* interest
"Little Steel formula," 477
Livingston, S. M., 869 *n.*
loan-deposits, 228, 232, 233, 246; *see* demand deposits
local expenditures, 674, 675, 677
lockout, 633, 637
long run, 318, 328, 391, 453
long-run supply, 378
  schedule, 331
losses, *see* profits
"low-consumption" economy, 124, 141
luxuries, 297

Macaulay, F. R., 490 *n.*
maintenance of membership, 629
margin of cultivation
  extensive, 451
  intensive, 451
margin requirements, 219, 583
marginal costs, 319-25, 399; *see* costs
marginal efficiency of capital, 128 *n.*
marginal product
  of capital, 434, 438, 439
  and income, 857
  of labor, 396-98
  relation to rent, 449-51
  relation to wages, 411, 415
  value of, 396, 400
marginal productivity theory, 394-405, 407, 409, 415, 416, 448
  criticisms of, 402-05
marginal revenue, 348, 349, 403
marginal utility, 293
Maritime Commission, 657, 658
Malthus, T. R., 406, 447 *n.*
"managed" money, 192
management security, 630
management, as a resource, 65
market, defined, 335
market demand, 301-311; *see also* demand
  schedule, 339
market economy, 274
market period, immediate, defined, 326
market price, 341
  under pure competition, 339-42
market ratio, 189
market regulation by government, 652, 653
market supply, 378; *see also* supply
  schedule defined, 329, 339
marketing agreements, 502
markets, relations between, 373-76
Marx, Karl, 851, 899 *ff.*, 921, 922, 925
Marxism, critique of, 907, 908
mass production, economies of, 589
mature economy thesis, 153-58
maximum prices, 277
Means, G. C., 554 *n.*
mediation, 636
member bank balances, 214, 234, 236, 239, 258, 269
member bank reserves, *see* member bank balances
member banks, 213
Mercantilism, 35-43
mergers, 594
merit rating, in unemployment compensation, 883, 884
Middle Ages, 29-35
milk marketing agreements, 508, 509
milk prices, control of, 508-10
Mill, John Stuart, 804
Miller-Tydings Amendment of 1937, 596, 601, 602, 654
Mills, F. C., 475 *n.*
minimum incomes, estimates of, 823, 824
minimum wage laws, 421, 422
  effect on labor supply, 860, 861
minority control of corporations, 555
mint par of exchange, 767; *see* foreign exchange
mint ratio, 189
Monetary Fund, *see* International Monetary Fund

monetary policy, 271
monetary standards, 187-97
money
  and capital, 70 *n.*, 392
  defined, 183, 184
  functions of, 180
  lawful, 198
  and prices, 248-59
  and socialism, 919
  see also currency *and* demand deposits
money economy, growth of, 37
money income, 15, 22
money supply, decrease of through debt retirement, 244-46
money system of the United States, 195, 196
money wages, 406, 487-89
monopolistic competition, 354-62, 379, 409
  defined, 338
monopoly, 164, 169, 273, 378, 498, 499
  bilateral, 373
  buyer, 371, 373; *see* monopsony
  defined, 338
  economies of, 592
  power, 540, 541
  price, 346-54
  privileges under mercantilism, 36, 41
monopsonistic competition, 372, 373
monopsony, 371
More, Sir Thomas, 895
Morris Plan banks, 205
mortgage bond, 565, 566; *see* corporation
motor-vehicle taxes, 705, 706
Moulton, H. G., 823, 824 *n.*
multilateral trade, 750, 751; *see* foreign trade
multiple basing points, 369
multiple unit banking system, 231-34
multiplier, the, 137, 138, 147
  and foreign trade balances, 807-09
Mun, 40

National Bank Act of 1863, 208
national bank note, 197, 209, 211, 259
National Banking System, 208
national banks, 197
National Bureau of Economic Research, 85
national currency, 197, 199
national debt, 196, 676, 727-33
  interest on, 730, 731
  burden of, 730-32
national defense expenditures, 676
National Farm Loan Associations, 558
national income, 15, 85 *ff.*, 92, 96-106
  defined, 89
  distributive shares, 97, 98
  and foreign trade, 805-12
  share of wages in, 643-47
  effect of war on, 147-50
national income payments, 90-92, 98
National Industrial Recovery Act of 1933, 519, 595, 611, 615, 634
  section 7a, 615, 634
National Labor Relations Act of 1935, 536, 634-37, 860, 861
National Labor Relations Board, 635-37
National Labor Union, 613
national output, 14, 85 *ff.*
  aggregate demand for, 91 *ff.*
  and capital goods, 94, 119
  composition of, 93-96
  and foreign trade, 805-12
  gross expenditure for, 91 *ff.*
  industrial sources of, 95, 96
  effects of war on, 147-50
National Recovery Administration, 510, 520, 521, 595
National Resources Committee, 71, 289, 586
national unions, 612 *ff.*
National War Labor Board, 629
nationalization of wealth, 853
"natural monopoly," 500, 588; *see* monopoly
natural resources, 53, 64, 65, 68, 69, 73; *see* conservation *and* land
"natural value," 345
nazi system, 912-13
Neal, A. C., 495 *n.*, 497 *n.*
necessities, 297
"needs test," 875
net national output, defined, 87; *see* national output
New Deal Administration, and labor organization, 615
New Deal spending, 724-26
New Economic Policy in Russia, 922-23
noncompeting groups, 858, 859
nonfinancial corporations, assets of, 553
non-participating income, 817-21, 851, 852, 873
nonprice competition, 354
nonunion labor, 620, 621

normal price, 343, 345-46
Norris-LaGuardia Act, 631, 633

occupational taxes, 704
Office of Price Administration, 513, 515, 516, 520
old age, cause of insecurity, 872, 873
old-age assistance, 877
old-age insurance, 865, 877, 885-88
old-age insurance for railroad workers, 888
old-age and survivors' insurance, 885-87
oligopoly, 379, 499
  defined, 338, 362, 363
  prices under, 364-71
oligopsony, 372, 373
open-market operations, 237
open-market policy, 267, 268
open shop, 629
operating capital, 112
opportunity and income, 858
opportunity cost, 419, 460, 461
output, level of, 128 *ff.*; *see* national output
  and profit expectations, 128 *ff.*
output, national, *see* national output
"overemployment," 151
overhead costs, 318-22, 385, 386
overissue, 194
"oversaving," 157, 841
"overspending," 841
overvalued money, 189, 190
Owen, Robert, 908
ownership and control, separation of, 554, 555

par of exchange, 767-69; *see* foreign exchange
parity prices, 491-94, 503; *see* agricultural prices
participating incomes, 817-21, 851, 852
participating preferred stock, 564; *see* corporation
partnership, 544, 545.
patents, 594
payments agreement, 798
payments, means of, 184; *see* money
Pearson, G. F., 256
pecuniary society, 182
perfect competition, 337 *n.*
perfectly elastic demand, 339; *see* elasticity of demand
perfect inelasticity of supply, 317; *see* elasticity of supply
personal distribution of income, 813 *ff.*; *see* income, distribution of
personal income, 90; *see* national income payments
personal income taxes, 98, 693; *see* income tax, individual
personal services
  demand for, 854-56
  and income inequality, 854 *ff.*
  supply of, 856-59
personal taxes, 693
Petty, Sir William, 64
physiocrats, 44 *n.*
picketing, 624
"Pittsburgh plus," 368, 369
planned economy
  contrasted with private enterprise, 523-25
  prices in, 522, 523
planning, economic, 913, 923-25
planning cost schedule of a firm, 323-25
Plato, 895
point system of rationing, 279
police power, 650
political economy, 3
poor relief, 876, 877
population
  ages of, 77, 78, 606, 607
  composition of, 606-10
  growth of, 53, 607, 608
  long-term changes of, 77-79
  and wages, 406, 407
  and world trade, 744, 745
Postal Savings System, 203, 206
preferences, scale of, 296; *see* consumption, theory of
preferred stock, 550, 564; *see* corporation
price competition, 354, 357, 358
  decline in, 366
price control, 482, 498 *ff.*
  in war-time, 278-79
price, defined, 273
price discrimination defined, 350
price economy, 274
price inflexibility and concentration, 495-98
price leadership, 367, 368, 595
price level, 249-51, 256, 263
price lines, 359
price movements, consequences of, 484

price structure, 473 *ff.*
price system, 274
  criticism of, 284-86
prices
  and allocation of consumer goods, 275-80
  and allocation of labor, 283; *see also* wages *and* income
  and allocation of resources, 380
  and composition of national output, 380; *see also* national output
  functions of, 377, 379
  and incomes, 383, 384
  in long-run under pure competition, 345, 346
  market, 339-42
  under monopolistic competition, 354-62
  under monopoly, 346-54
  under oligopoly, 364-71
  under pure competition, 338-46
  and production, 280-84
  of productive services, 382, 386-89
  in Russia, 926
  under socialism, 523, 919
primary deposits, 228, 232, 233, 246; *see* demand deposits
private debt, contrasted with public, 673
private enterprise
  functions of, 530-34
  contrasted with planned economy, 523-25
private finance, contrasted with public finance, 670-74
private investments, compared with government expenditures, 126
producer decisions, 109-19
producer goods defined, 9, 70; *see* capital goods
product differentiation, 356, 365
product variation, 358-59
production, 14-18
  agents of, 394
  allocation of, 59; *see also* resources, allocation of
  and consumer expenditures, 845, 846
  factors of, 64 *ff.*, 381 *ff.*
  and investment, 122
  large scale, 589, 590
  levels of, 59; *see also* national output
  under Mercantilism, 38-40
  in Middle Ages, 32, 33
  and prices, 280-84
production (Cont.)
  roundabout, 430
  specialization of, 61-63
  stages of, 61, 62
  and wants, 60, 61
Production Credit Associations, 558
productive capacity, 80-83
productive income, 98
productive services
  classification of, 384, 386
  defined, 381
  demand for, 399; *see* derived demand
  equilibrium price of, 402
  optimum combination of, 400, 401
  prices of, 382, 386, 388, 389
productivity
  diminishing, *see* diminishing returns
  marginal, *see* diminishing returns
productivity of public expenditures, 682-84
professional workers, 76
profit-and-loss statement, 572
profit expectations, 109, 110, 472
  in the 1930's, 145 *n.*
  and income, 833
  and national output, 128 *ff.*
profit motive, 25, 26
profits, 393, 459 *ff.*
  and allocation of resources, 467-68
  business, 460
  implicit profits, 460
  and innovations, 467
  and level of output, 472
  and market imperfections, 465, 466
  and monopoly, 467
  pure, 460
  under pure competition, 461, 462
  and risk, 468-71
  and uncertainty, 462 *ff.*
progress, the conditions of, 50-56
progressive taxation, 697, 713
  and income distribution, 838, 840, 844
propensity to consume, 100, 124, 125, 141, 832, 833, 839-41
  average, 136
  changes during war, 148, 149
  marginal, 136, 137
propensity to invest, 109
propensity to save, 100, 109, 832, 833
  changes during war, 148, 149
property tax, 702-04
  incidence of, 721

proportional taxation, 712
proportionate distribution of income, 822-30; *see also* income, distribution of
proprietors and managers, nonfarm as gainful workers, 76
proprietorship, 543, 544
prospectus, 581
protectionism, arguments for, 789-93; *see* protective tariff
protective tariff, 788 *ff.*
Proudhon, 851
proxy, 548
public budgets, 688-90
public debt, contrasted with private, 673; *see* national debt
public decisions, 106, 107
public expenditures
  allocation of, 684-86
  determination of, 684-86
  growth of, 674, 675, 678-80
  productivity of, 682-84; *see* government expenditure
public finance, 669 *ff.*
  contrasted with private finance, 670-74
public revenues, 687, 688
public sector of the economy, 669 *ff.*
public works, 865, 877, 878
Public Utility Holding Company Act of 1935, 552, 553
public utility regulation, 655-61
  rate regulation, 759-61
purchasing power, 183, 186, 187
purchasing power parity, 772, 773, 775; *see* foreign exchange
pure competition, 371, 378
  defined, 337, 338
  prices, 338-46
pure interest, 442
pure profits, 460 *ff.*

quality competition, 354, 358-60
quantity theory of money, 249-52, 256, 262
quasi-rent, 391, 453-56
quotas, import, 795, 796

railroad workers, social insurance of, 888
railroads, regulation of, 656, 657
Railway Labor Act, 634
rate of exchange in international finance, 766; *see* foreign exchange
rationing, 275, 276, 278, 279, 516, 517
  point system, 279
real expenditures, 680-82
real income, 15, 22, 102
real output, 92, 103, 145, 147
real wages, 406, 488, 489
receipts, prospective, and investment, 111, 112, 114-16
Reciprocal Trade Agreements Act, 793, 794
Reconstruction Finance Corporation, 206, 216
rediscount policy, 265-67
  rate, 265-68
rediscounting, 215, 216
registration of securities, 581, 582; *see* Securities Exchange Commission
regressive tax, 714
regulation of industry, 585 *ff.*; *see* government control
Regulation W, 219
relative prices, movements of, 474-76
relief, 876-78
rent, 391, 446 *ff.*
  and cost, 456, 457
  defined, 446
  function of, 457
resale price maintenance, 601, 602, 654
reserve bank credit, 236-41, 247, 258; *see* Federal reserve banks
reserve city banks, 210, 214
reserve plan, 100 per cent, 263, 271
reserve requirements, alteration of, 268
reserves
  of Federal reserve banks, 234, 235
  gold certificates as, 195
  of member banks, 214, 240, 241, 268
  under national banking system, 210, 211; *see also* currency
resources, 10, 63-66
  allocation of, 530-34
  allocation under monopoly, 353
  allocation and profits, 380
  allocation by prices, 380
  and income distribution, 831 *ff.*
  and international trade, 739, 740
  types of, 57 *ff.*
restraint of trade, 597
retained earnings, 90, 102, 569
revenue, federal tax, 695
Ricardo, David, 447, 804, 868
rigid prices, 366

risk, 65, 66, 469, 470
and investment, 110, 111, 117, 163, 164
and profit expectations, 110, 111, 468-71
rival supply, 333, 376
Robinson-Patman Act of 1936, 596, 601, 654
Rochdale co-operatives, 556
roundabout production, 430
"rule of reason," 595, 598
Rural Electrification Administration, 558
Russia, prices in, 926

sales schedule of a monopolist, 348
sales taxation, 706, 707
savings, 98-100
and capital, 10, 429 *ff.*
and investment, 121, 122, 433-35
and liquidity preference, 435 *ff.*
personal, 98, 100
ratio, 100
see propensity to save
savings
and time preference, 431-35
savings banks, 204
savings bonds, 727; *see* national debt
savings deposits, 185, 204; *see* time deposits
scale of preferences, 296
scarcity, 9
relation to costs, 12
relation to supply, 313
relation to wants, 3
Schumpeter, J. A., 165 *n.*, 168 *n.*, 171
"scientific" tariff, 791, 792
seasonal movements of prices, 485, 486
seasonal unemployment, 868
secular unemployment, 870
Securities Act of 1933, 581
Securities and Exchange Act of 1934, 219, 582, 583
Securities and Exchange Commission, 552, 581-85
securities transactions in balance of payments, 754-59
security issues, control of, 580-82
security, 174-77
social, *see* social insurance
seigniorage, 200, 688
self-interest, 531; *see* profit motive *and* theory of consumption
self-sufficiency, 792, 793
semiskilled workers, 76
seniority, 628
service industries, 66 *n.*, 93
services
contrasted with commodities, 14
in international transactions, 749, 750, 754, 755, 757
as part of national output, 87
Shaw, George Bernard, 897, 909
shareholders, *see* corporation, ownership of
shares, 550; *see* corporation stock
Sherman Anti-Trust Act of 1890, 363, 537, 588, 596-600, 602, 632, 653, 654
Sherman Silver Purchase Act, 260
shifting and incidence of taxation, 717-22
short run, 391, 453
defined, 318, 327
short-run prices under pure competition, 342
short-run supply, 378
schedule, 327, 330
silver, abandonment of, 259
silver certificates, 197, 199, 261
silver dollars, 197, 199, 200
silver, price of, 200 *n.*, 259-61
free and unlimited coinage, 259
remonetization of, 260, 261
Silver Purchase Act of 1934, 260, 261
sit-down strikes, 623
size of firm, 323-25
skilled workers, 76
"slow-down," 623
Smith, Adam, 45, 46, 58, 109, 165, 494, 498, 542, 548, 710, 741, 742, 804, 868
social democrats, 910, 911, 913
social insurance, 865-67, 874, 876
costs of, 889-92
and purchasing power, 890-92
Social Security Act, 875, 881 *ff.*, 889
Social Security Administration, 872 *n.*, 882
social security taxes, 91
socialism, 668, 894 *ff.*
allocation of resources under, 919-21
in Britain, 928, 929
contrasted with capitalism, 929-32
criticism of, 913, 914
Fabian, 909-10
planning under, 523
prices under, 523, 919
in Russia, 921-28
and scarcity, 917
"scientific," 901

socialism (Cont.)
utopian, 908, 909
see also Marx
Soil Conservation and Domestic Allotment Act, 502
Soviet Russia, 28, 55 *n.*, 921-28
"soviets," 906
special assessments, 688
specific duty, 788; *see* tariffs
spending ratio, 100; *see* propensity to consume
stabilization
monetary policy of, 266, 270, 271
monetary proposals, 173-77, 262-64
Stabilization Fund, 195, 773-75
stages of production, 61, 62
standard of living, 57; *see* wants
relative to wages, 406, 407
state banks, 197
state expenditures, 674, 675, 677
state and local taxes, 702
state trading, 802, 803
Stevenson Plan, 505
stock market, 219
stock, preferred, 550, 564; *see* corporation
stockholder, *see* corporation, ownership of
strikes, 622-24, 637
structural unemployment, 870
subsistence theory of wages, 406, 407
subsistence wage, 41, 42
substitutes, 297, 306, 311, 347
price relations between, 374, 376
supply, 313 *ff.*
aggregate and particular contrasted, 288, 334
changes in, 314, 315
cost relations, 326-32
defined, 313
elasticity of, 316, 332
interrelations of, 332, 333
perfect inelasticity of, 317
schedule, 313 *ff.*, 378
schedule of individual firm, 321
schedule, long-run, 331
schedule, short-run, 327, 342-44
supply of foreign exchange, 767-80; *see* foreign exchange
supply of labor, 414, 415, 605 *ff.*
supply of money, 254
control by Federal Reserve, 268-71
relation to rate of interest, 436-38
supply of savings, 436, 437
support prices, 503; *see* agricultural prices
surpluses, relation to rent, 456, 457
surplus value, 901-04, 907, 908
sympathy strikes, 623
syndicalism, 910, 911
syndicate in investment banking, 580

Taft, Philip, 618 *n.*, 619 *n.*
Taft-Hartley Act of 1947, 537, 625 *n.*, 636, 637
Tarasov, Helen, 693 *n.*
tariffs, 788 *ff.*
Tawney, R. H., 31, 33 *n.*
tax collections, relation to money system, 238
tax, individual income, 693-99
tax policy, 709 *ff.*
tax revenues, federal, state, and local, 708, 709
tax system, requirements of a, 710-16
of the United States, 691 *ff.*
taxable income, 696
taxation
ability to pay principle of, 711-16
benefit principle of, 711, 712
burden of, 691 *ff.*
canons of, 710
economic effects of, 716-22
effects upon consumption, 175, 716
effects upon investment, 163
equity of, 711-16
progressive, 697, 713
proportional, 712
regressive, 714
shifting and incidence of, 717-22
technological unemployment, 868
technology, 79
improvements in, 53, 54
and scale of production, 329, 332
and wages, 419-21
Temporary National Economic Committee, 585
Tennessee Valley Authority, 665, 666
Terborgh, George, 158 *n.*, 161 *n.*
Thomas, Woodlief, 184 *n.*
Thorp, W. L., 496 *n.*
time-deposit, defined, 185
time preference-productivity theory of interest, 431-35
totalitarianism, 912; *see* collectivism, communism, fascism

tourist expenditures, in balance of payments, 756 *n.*
trade associations, 534, 594, 595
transfer expenditures
  effects on consumption, 680-82
  and incomes, 817
  effects on investment, 681, 682
  effects on savings, 681, 682
transfer payments, 90 *n.*, 98, 143; *see* transfer expenditures
Treasury, United States, 234; *see* Treasury balances *and* Treasury cash and deposits
Treasury balances with Federal reserve banks, 238-40
Treasury bills, 728; *see* national debt
Treasury bonds, 727; *see* national debt
Treasury cash and deposits, relation to member bank balances, 236, 238-40, 258
Treasury notes, 727; *see* national debt
Treasury notes of 1890, 197, 199
Treasury policy, relation to Federal Reserve Policy, 270
triangular exchange, 765; *see* foreign trade *and* international finance
Trotsky, 905, 911
trust, business, 593, 594
Trust Indenture Act of 1939, 584
"trust-busting," 598, 599
"Truth in Securities Act," 581

uncertainty, and profits, 462 *ff.*
undervalued money, 189, 190
underwriting of securities, 580
underwriting syndicate, 545
undistributed earnings, 90, 102, 569
undistributed profits, taxation of, 699, 700
unemployment, 76, 83, 147, 867-71
  cyclical, 869, 870
  frictional, 870, 871
  insurance, *see* unemployment compensation
  irreducible minimum of, 871
  relief, 865
  in Russia, 927, 928
  seasonal, 868
  secular, 870
  technological, 868
unemployment compensation, 874, 881-85
Unemployment Compensation Trust Fund, 883
unfair labor practices, 635, 636
unfavorable balance of trade, 253, 749; *see* international trade
unilateral transfers, 757-59; *see* balance of international payments
union contracts, 627-30
union discipline, 619, 620
union finances, 619
union jurisdiction, 615
union recognition, 628, 629
unions, 534-38, 611 *ff.*
  growth of, 612-16, 640, 642
  and labor supply, 860, 861
  and national output, 644-46
  political activities of, 625, 626
  and production, 645
  and wages, 421-24, 644-46
  welfare activities of, 621
unit banking, 222
United States Forest Service, 663
United States notes, *see* greenbacks
United States Tariff Commission, 791
unlimited liability, 545
unskilled workers, 76
urban rents, 452, 453
urbanization and old-age problem, 872, 873
use tax, 707
utility, 62, 292
  marginal, 293
  total, 293
utopian socialism, 908, 909

valuation of corporate assets, 576
valuation of public utility property, 660
value of the marginal product, 396
  relation to cost, 400
  relation to price, 400
  see marginal product
variable costs, 318-22
velocity of circulation, 250-52
vertical combination, 591
voluntary arbitration, 638

wage bargaining, 622, 626-30
wage indexes, 486-89
wage rates, 486, 487
wage theories, 406-09
wages
  bargaining theory of, 409
  and cost of living, 423
  defined, 405

wages (Cont.)
differences in, 416-19
and employment, 424-28
and inventions, 419-21
marginal productivity theory of, 394-405, 407, 409, 415, 416, 448
and national income, 412, 643-47
and population, 406, 407
and prices, 427, 428
and purchasing power, 425-27
subsistence theory of, 406, 407
and supply of labor, 414, 415
and unemployment, 424-28
and unions, 421-24
wages-fund theory of, 408
wages-fund theory, 408
wages of management, 405
waiting, as a cost, 11, 12
wants, 3-8
in medieval society, 31, 32
and production, 60, 61
and scarcities, 313
war bonds, 244; *see* savings bonds *and* national debt
war, extended use of resources in, 82
war financing and prices, 512, 513
War Food Administration, 516
War Labor Board, 477, 510
war materials as a component of national output, 94
War Production Board, 516
Warburton, C., 823, 824 *n.*
Warren, G. F., 256, 257
wartime price control, 511-19
waste in public expenditure, 683, 684
Watson, Donald S., 724 *n.*, 726 *n.*
wealth, distribution of, 852-54
and income inequality, 852-54
Webb, Sidney and Beatrice, 909
Webb-Pomerene Act, 801
Weber, Max, 33 *n.*
weighted index number, 480
Wells, H. G., 909
West, David, 447 *n.*
wholesale prices, 482-84
in world wars, 514
Wilcox, Clair, 338 *n.*, 499 *n.*
"wildcat" banks, 208
Williams, F. E., 75 *n.*
Wingfield, B. Magruder, 221 *n.*
women, in labor supply, 76, 77
work relief, 865, 877, 878
working capital, 112, 562
working capital ratio, 577
workmen's compensation, 865, 874, 879-81
world trade, and population, 744, 745

yellow dog contract, 633, 634

zone price systems, 368, 369